WO9505389

Publication Title:

SEQUENCE-SPECIFIC GLYCOCONJUGATE TRANSCRIPTIONAL ANTAGONISTS

Abstract:

Abstract of WO9505389

The invention provides glycoconjugates which bind polynucleotides in a sequence-selective manner and/or preferentially displace or inhibit binding of transcription factors to their recognition sites on DNA. The DNA-binding glycoconjugates of the invention, exemplified by calicheamicin-MG (as shown in the figure), are used as selective transcriptional antagonists, among other uses.

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C07H 13/12, C12Q 1/68

A1

(11) International Publication Number: WO 95/05389

(43) International Publication Date: 23 February 1995 (23.02.95)

(21) International Application Number: PCT/US94/09123

(22) International Filing Date: 15 August 1994 (15.08.94)

(30) Priority Data:
08/109,271 18 August 1993 (18.08.93) US

(71) Applicants: THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY [US/US]; Stanford, CA 94305 (US). YALE UNIVERSITY [US/US]; Suite 401, 246 Church Street, New Haven, CT 06510 (US). THE PRESIDENT AND FELLOWS OF HARVARD COLLEGE [US/US]; University Place, 4th floor South, 124 Mt. Aubern Street, Cambridge, MA 02138 (US).

(72) Inventors: HO, Steffan, N.; 736 Partridge Avenue, Menlo Park, CA 94025 (US). SCHREIBER, Stuart, L.; 434 Marlborough Street, Boston, MA 02115 (US). DANISHEFSKY, Samuel, J.; 57 Stevenson Road, New Haven, CT 06515 (US). CRABTREE, Gerald, R.; 7 Durham Road, Woodside, CA 94062 (US).

(74) Agents: DUNN, Tracy, J. et al.; Townsend and Townsend Khourie and Crew, 20th floor, One Market Plaza, Steuart Street Tower, San Francisco, CA 94105 (US).

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

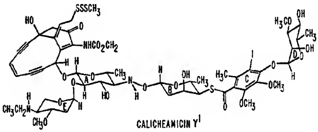
Published

With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

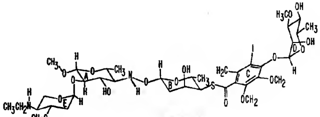
(54) Title: SEQUENCE-SPECIFIC GLYCOCONJUGATE TRANSCRIPTIONAL ANTAGONISTS

(57) Abstract

The invention provides glycoconjugates which bind polynucleotides in a sequence-selective manner and/or preferentially displace or inhibit binding of transcription factors to their recognition sites on DNA. The DNA-binding glycoconjugates of the invention, exemplified by calicheamicin-MG (as shown in the figure), are used as selective transcriptional antagonists, among other uses.



CALICHEAMICIN γ[1]



CALICHEAMICIN - MG

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgystan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SI	Slovenia
CI	Côte d'Ivoire	KZ	Kazakhstan	SK	Slovakia
CM	Cameroon	LI	Liechtenstein	SN	Senegal
CN	China	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
CZ	Czech Republic	LV	Latvia	TJ	Tajikistan
DE	Germany	MC	Monaco	TT	Trinidad and Tobago
DK	Denmark	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	US	United States of America
FI	Finland	ML	Mali	UZ	Uzbekistan
FR	France	MN	Mongolia	VN	Viet Nam
GA	Gabon				

SEQUENCE-SPECIFIC GLYCOCONJUGATE TRANSCRIPTIONAL ANTAGONISTS

This invention was made in the course of work supported by the U.S. Government and Howard Hughes Medical Institute, which may have certain rights in this invention.

FIELD OF THE INVENTION

The invention provides compositions and methods for modulating transcriptional activity of specific genes in eukaryotic cells by selectively inhibiting binding interactions between DNA-binding proteins (e.g., transcription factors) and their DNA recognition sites. The invention encompasses novel DNA-binding glycoconjugates, pharmaceutical and biochemical reagent compositions thereof, methods for modulating transcriptional activity of predetermined gene sequences, methods for treating or preventing pathological conditions by modulating gene expression, methods for suppressing T cell activation to provide therapeutic or prophylactic immunosuppression, and methods for identifying novel glycoconjugates that are sequence-specific transcriptional antagonists.

BACKGROUND

Many pathological conditions and genetic diseases, including but not limited to neoplasia, autoimmune processes, infectious diseases, and the like could be treated or prevented if there were methods and compositions to facilitate the directed modulation of the transcriptional activity of a predetermined gene (or genes) in cells. Several approaches to modulating gene expression have been investigated, with variable success.

Transgenesis, or transferring an exogenous expression gene cassette into a cell, has been used for placing an exogenous gene linked to a predetermined transcriptional control sequence into a cell or a non-human animal embryo. Transgenesis has been used in many applications, including human <u>in vivo</u> and <u>ex vivo</u> gene therapy, making transgenic animals for commercial and academic research, and the like. Homologous gene targeting has

also been employed to make so-called "knockout" animals, having functionally disrupted predetermined genes (Jasin and Berg (1988) Genes and Development 2: 1353; Doetschman et al. (1988) Proc. Natl. Acad. Sci. (U.S.A.) 85: 8583; Dorini et al. (1989) Science 243: 1357; Itzhaki and Porter (1991) Nucleic Acids Res. 19: 3835; Valancius and Smithies (1991) Mol. Cell. Biol. 11: 1402; Koller et al. (1991) Proc. Natl. Acad. Sci. (U.S.A.) 88: 10730 and Snouwaert et al. (1992) Science 257: 1083). However, inactivating gene function by homologous gene targeting is difficult and generally leads to a complete ablation of gene function.

Antisense oligonucleotides have also been proposed as "code blockers" that will inhibit the expression of specific genes on the basis of sequence-specific binding, either to inhibit translation, transcription, or both (Ching et al. (1989) Proc. Natl. Acad. Sci. U.S.A. 86: 10006; Broder et al. (1990) Ann. Int. Med. 113: 604; Loreau et al. (1990) FEBS Letters 274: 53; Holcenberg et al. WO91/11535; Cooney et al. (1988) Science 241: 456; Antisense RNA and DNA, (1988), D.A. Melton, Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY). One mechanism by which sense or antisense oligonucleotides may function is via formation of triplex or quadruplex DNA structures in a corresponding chromosomal gene locus (Cheng et al. (1988) J. Biol. Chem. 263: 15110; Ferrin and Camerini-Otero (1991) Science 354: 1494; Ramdas et al. (1989) J. Biol. Chem. 264: 17395; Strobel et al. (1991) Science 254: 1639; Rigas et al. (1986) Proc. Natl. Acad. Sci. (U.S.A.) 83: 9591; Camerini-Otero et al. U.S. 7,611,268; WO93/05178; Beal PA and Dervan PB (1991) Science 251: 1360). Antisense polynucleotides theoretically have the advantage of producing reversible effects and of producing titratable dose-response relationships, albeit there is little experimental evidence demonstrating these properties.

Both approaches to controlling gene expression, transgenesis/homologous recombination and antisense/triplex require transferring exogenous polynucleotide sequences into cells, which is problematic, especially for pharmaceutical administration. Further, both approaches tend to produce a long-

lived effect and are poorly titratable (i.e., they tend to produce an "all-or-none" modulation of gene activity). Perhaps more importantly, both approaches have met with limited success for producing desired modulation of specific gene activity in vivo.

Some contemporary small molecule pharmaceuticals (i.e., molecular weight less than about 1,000-3,000 daltons) act to modulate gene expression of particular genes or gene subsets, and usually have advantageous pharmacological properties compared to oligonucleotides. However, most such small molecules usually act in an indirect fashion that is not sequence-specific (e.g., they act by inhibiting an "upstream" signal transduction pathway that does not consist of transcriptional complexes). For example, glucocorticoids and estrogen/progesterone contraceptives modulate steroid responsive gene expression by interacting with proteinaceous steroid hormone receptors, and do not themselves form direct sequence-specific interactions with predetermined target gene sequences. Frequently, such small molecule pharmaceuticals possess undesirable side effects or affect a broad spectrum of expressed gene sequences where only a specific effect on the expression of a particular gene is desired.

Thus, there exists a need in the art for sequence-selective DNA-binding agents that can be used to modulate expression of predetermined genes and which can be formulated for facile and efficacious delivery into cells and organisms. The present invention fulfills these and other needs.

The references discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the inventors are not entitled to antedate such disclosure by virtue of prior invention. All publications and patent applications herein are incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

SUMMARY OF THE INVENTION

The present invention provides several novel methods and compositions for modulating the transcriptional activity of genes that are under the transcriptional regulation of predetermined regulatory sequences to which transcription factors, accessory proteins, or other DNA-binding proteins bind. The present invention provides novel glycoconjugate DNA ligands which bind to a predetermined cis-acting DNA sequence and inhibit or disrupt binding of a DNA-binding protein to said predetermined cis-acting DNA sequence, thereby producing a modulation in the transcription rate of a structural gene(s) under the transcriptional control of the predetermined cis-acting DNA sequence. Typically, the predetermined cis-acting DNA sequence is a binding site for a transcription factor, such as a eukaryotic transcription factor, often an inducible eukaryotic transcription factor (e.g., nuclear activator of activated T cells, or NFAT), or other site at which a polypeptide having transcriptional regulatory activity binds. By interacting with the cis-acting DNA sequence, the glycoconjugate DNA ligand displaces, disrupts, or inhibits the binding of a DNA-binding transcriptional regulatory protein to the cis-acting sequence and interferes with (i.e., inhibits) the activity of the DNA-binding transcriptional regulatory protein, whether the DNA-binding transcriptional regulatory protein is a positive or negative regulator of transcription at the locus. Advantageously, these novel glycoconjugate DNA ligands reversibly bind to the DNA recognition sequences in a sequence-specific or sequence-selective manner and do not form reactive species (e.g., radicals) that damage or cleave the DNA. Such glycoconjugate DNA ligands can be used for various applications, including but not limited to their use as pharmaceuticals for effecting therapeutic modulation of the transcription of specific genes.

The glycoconjugate DNA ligands of the present invention bind to DNA, preferably double-stranded DNA in the B helical form, in a sequence-selective manner such that the glycoconjugate DNA ligand preferentially binds to a sequence motif in chromosomal DNA of about at least four nucleotides in length,

thus forming a liganded DNA recognition site complex. In one embodiment, the glycoconjugate DNA-ligand binds to a predetermined DNA recognition site at which a DNA-binding protein is presently bound (i.e., an occupied recognition site) and displaces the bound DNA-binding protein from the recognition site, thus forming a liganded DNA recognition site complex which does not contain the DNA-binding protein. In another embodiment, the glycoconjugate DNA ligand binds to a predetermined recognition site to which a DNA-binding protein is presently bound without dissociating the DNA-binding protein from the site, thus forming a liganded DNA recognition site complex which contains the DNA-binding protein, but wherein the biological activity of the DNA-binding protein at the predetermined recognition site is inhibited or altered. In a further embodiment, a glycoconjugate DNA ligand preferentially binds to an unoccupied predetermined recognition site and inhibits binding of the DNA-binding protein to its recognition site sequence.

The present invention also provides compositions containing glycoconjugate DNA ligands that can serve as reagents to displace a bound DNA-binding protein from its recognition site and/or inhibit binding of a DNA-binding protein to its recognition site sequence. These reagents comprise an effective concentration of a glycoconjugate DNA ligand and will be used in a variety of commercial and research applications, such as eluting DNA-binding proteins from a predetermined DNA recognition sequence, such as in eluting specifically bound NFAT protein(s) from a polynucleotide or oligonucleotide affinity column containing a NFAT polynucleotide recognition sequence. In one variation, these reagents are used for preparative scale commercial purification of NFAT from cell or nuclear extracts; such purified NFAT will itself find use as a commercial research or diagnostic reagent. In some embodiments, compositions comprising a therapeutically or prophylactically effective dosage of a glycoconjugate DNA ligand are formulated for administration to a human or veterinary patient for treating a disease condition by modulating the transcription of predetermined genes for therapeutic benefit. In one variation, a glycoconjugate DNA

ligand, such as an alkyl glycoside derivative of calicheamicin γ_1^I (e.g., CLM-MG) is administered to a patient to inhibit binding of the NFAT transcription factor to its recognition sequence(s) and, for example, to inhibit T cell activation and effect immunosuppression to prevent tissue graft rejection reactions.

The invention also provides methods for identifying novel sequence-selective glycoconjugate DNA ligands that bind preferentially to a predetermined DNA recognition sequence to which a transcription factor can bind. These methods generally comprise the steps of (1) synthesizing a glycoconjugate DNA ligand having a carbohydrate recognition element comprising a mixed polysaccharide/aryl substituent having the backbone structure of calicheamicin $\gamma 1^I$, with one or more substituent modifications or substitutions (e.g., replacing the iodo substituent on the thiobenzoate ring with a bromo, chloro, or methyl substituent) and lacking the reactive diyne-ene ring structure, optionally also in covalent linkage to one or more non-interfering substituents replacing the diyne-ene substituent position, (2) contacting the glycoconjugate DNA ligand with a polynucleotide having a predetermined recognition sequence to which a predetermined sequence-specific DNA-binding protein (e.g., transcription factor) binds in the presence of said sequence-specific DNA-binding protein under binding conditions (e.g., an NFAT recognition site), (3) determining the ability of the glycoconjugate DNA ligand to inhibit the binding of the predetermined sequence-specific DNA-binding protein to the predetermined DNA recognition sequence, and (4) identifying glycoconjugate DNA ligands which inhibit said protein binding to the predetermined DNA sequence as being a sequence-selective glycoconjugate DNA ligand. Typically, sequence-selectivity is demonstrated by including in the binding reaction additional species of DNA binding proteins (i.e., different than said predetermined DNA binding protein), termed "alternative DNA binding proteins", and additional polynucleotides having different recognition DNA sequences (i.e., different than said predetermined DNA recognition sequence), termed "alternative DNA

recognition sequences" which bind to said alternative DNA binding proteins, and determining that the glycoconjugate DNA ligand preferentially inhibits binding of the predetermined DNA binding protein to its predetermined DNA recognition sequence as compared with the alternative species of DNA binding proteins and their cognate alternative DNA recognition sequences. Usually, alternative DNA binding proteins and alternative DNA recognition sequences are transcription factors and transcription factor recognition sequences, respectively, that are distinct from the predetermined DNA-binding protein and the predetermined DNA recognition sequence(s). Thus, if the predetermined DNA-binding protein is NFAT and the predetermined DNA recognition sequence is an NFAT recognition site (e.g., 5'-AAGGAGGAAAAACTGTTTCAT-3'), examples of alternative DNA-binding proteins are Sp1 and AP1 and their cognate DNA recognition sequences 5'-GGGGCGGGGC-3' and 5'-GTGACTCAGCGCG-3', among others.

In a variation of the above-described general method for identifying novel sequence-selective glycoconjugate DNA ligands that bind preferentially tò a specific DNA sequence, said ligands are identified by screening a library of glycogonjugate DNA ligands with a predetermined DNA sequence. Alternatively, DNA sequences that bind to de novo synthesized glycoconjugate compounds that lack a predetermined DNA recognition site can be identified by screening a library of DNA sequences with a unique glycoconjugate ligand.

Glycoconjugate DNA ligand-mediated inhibition of binding of a transcription factor to its recognition site(s) may be conveniently assayed using a transcription assay, such as an _in vitro_ transcription assay or a transient (or stable) transfection assay employing a reporter gene operably linked to a transcriptional regulatory sequence comprising the predetermined transcription factor recognition site(s). A glycoconjugate DNA ligand is introduced into or administered to a transcription assay wherein sequence-specific transcriptional regulation of a reporter gene is effected by a predetermined cis-acting operably linked transcription factor recognition site. Application of a glycoconjugate DNA ligand that preferentially

binds to the predetermined transcription factor recognition site and inhibits the transcriptional modulation conferred by the transcription factor recognition site on the transcription of the cis-linked reporter gene will result in altered transcription (e.g., increased or decreased as compared to a control assay reaction lacking the glycoconjugate DNA ligand) of the reporter gene. The ability of a glycoconjugate DNA ligand to inhibit the transcriptional effect produced in the control assay by binding of the transcription factor to its recognition site(s) in the predetermined transcriptional regulatory sequence will indicate that the sequence-selective glycoconjugate DNA ligand and/or its related structural congeners can be used as a modifier of specific gene expression for genes or other expressible polynucleotides under a transcriptional influence of the predetermined transcriptional regulatory sequence.

The activity of a glycoconjugate DNA ligand as a modifier of specific gene expression may be employed for therapeutic benefit to treat diseases wherein modification of the expression of specific genes or gene subsets can reduce or arrest the pathological condition (e.g., by immune modulation, reversal of neoplastic transformation, inhibition of metastatic phenotypes, for gene therapy, and the like). For therapeutic or prophylactic treatment of a disease condition, a therapeutic or prophylactic dosage of a sequence-selective glycoconjugate DNA ligand of the invention is administered to a patient; such a therapeutic or prophylactic dosage will be sufficient, alone or in combination with another pharmaceutical agent, to alter the transcription rate and/or net transcription of specific genes under the transcriptional influence of the predetermined DNA recognition sequence to which the glycoconjugate DNA ligand preferentially binds.

The invention also provides therapeutic and/or prophylactic compositions of sequence-selective glycoconjugate DNA ligands for use as pharmaceuticals or reagents; such compositions comprise an effective amount or concentration of one or more sequence-selective glycoconjugate DNA ligands.

In one embodiment, a glycoconjugate DNA ligand having

a calicheamicin-related structure lacking an enediyne core, calicheamicin MG, is provided for inhibiting the transcription of at least one structural gene under the cis-acting transcriptional influence of a DNA recognition site sequence for the transcription factor NFAT. The specific inhibition of one or more NFAT-dependent genes may be advantageously accomplished by administration of the glycoconjugate DNA ligand to produce a modulation of immune function, such as an inhibition of T lymphocyte activation (e.g., IL-2 expression) for reducing graft rejection reactions, among other uses.

BRIEF DESCRIPTION OF THE FIGURES

Fig. 1 shows the structures of the enediyne antibiotic calicheamicin $\gamma 1^I$ and the glycoconjugate DNA ligand calicheamicin-MG, which represents the methyl glycoside derivative of calicheamicin $\gamma 1^I$.

Fig. 2 shows the transcription factor recognition sequences used. NFAT site is the human IL-2 enhancer sequence -286 to -267; AP1 site is derived from the human metallothionein enhancer; NFκB site is derived from the murine immunoglobulin kappa light chain enhancer; Oct1 site is the human IL-2 enhancer sequence -85 to -65; Oct1/octamer-associated protein (OAP) site is the human IL-2 enhancer sequence -101 to -65; HNF-1 site is the human β-fibrinogen enhancer sequence -103 to -76; Sp1 site is the consensus Sp1 binding site. Sequences in lower case represent potential calicheamicin binding site(s).

Fig. 3 shows sequence-selective inhibition of DNA-protein complex formation by calicheamicin MG (CLM-MG). Lane 1, buffer control; lanes 2-5, CLM-MG added to a final concentration of 100 μM, 32 μM, 10 μM, and 3.2 μM, respectively. NFAT, AP1, and NFkB activities are from TAg Jurkat nuclear extracts; Oct1, Oct1/OAP, and Sp1 activities are from HeLa nuclear extracts; HNF1 activity is from MDCK cell nuclear extract. Identical results were obtained for Oct1, Oct1/OAP, and Sp1 activities from TAg Jurkat cells, and for AP1 and NFkB activities from HeLa cells. NS indicates DNA binding activity non-saturable by a 100-fold excess of cold probe oligonucleotide.

Fig. 4 shows quantitation of gel shift data for NFAT (solid circles, n=4), AP1 (solid squares, n=3), and HNF1 (open squares, n=3). Results are presented as percent inhibition of the DNA binding activity measured in the absence of CLM-MG and represent mean +/- standard error of the mean.

Fig. 5 shows representative gel shift data demonstrating that alteration of the CLM recognition element within the NFAT binding site alters sensitivity of NFAT binding to inhibition by CLM-MG. The NFAT DNA probe sequences used are shown (in part) to the left of each gel shift panel. The upper sequences represent the wild-type NFAT probe sequence. The middle and lower sequences contain sequence changes (indicated in bold) that alter the CLM-MG binding site. DNA-protein complexes were allowed to form at room temperature for 60 minutes and transferred to 4°C. After 15 minutes, control buffer (lane 1), CLM-MG at approximately 100 μM, 32 μM, 10 μM (lanes 2-4, respectively), or 100-fold excess cold probe (lane 5) was added. After one minute, the samples were electrophoresed.

Fig. 6 represents a quantitation of the gel shift data exemplified in Fig. 5. Open squares represent wild-type NFAT probe; closed squares represent an A insertion (n=4); closed circles represent a G to T substitution (n=4). Points represent mean; error bars indicate standard error of the mean. The nucleotide sequences are those shown in Fig. 5.

Fig. 7 shows sequence-specific inhibition of transcription *in vivo* by CLM-MG. (Top panel) TAg Jurkat cells were transfected with the indicated secreted alkaline phosphatase reporter DNA. Approximately 24 hours later, the cells were incubated with 200 μM CLM-MG or buffer control for 15 minutes. The cells were then stimulated with ionomycin and TPA, resulting in a final concentration of 100 μM CLM-MG. Secreted alkaline phosphatase was measured after 12-16 hours. Results are expressed as percentage of reporter gene activity stimulated by ionomycin plus TPA in absence of CLM-MG and represent the mean +/- standard error of the mean. (Bottom panel) The percentage of inhibition of reporter gene expression dependent on various transcription factors was determined as above except the

concentration of CLM-MG was varied from 0 to 250 μM to demonstrate concentration-dependence. NFAT-SX (open circle), AP1-SX (closed circle), NFkB-SX (open square), or Oct1-SX (closed square).

DEFINITIONS

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods and materials are described. For purposes of the present invention, the following terms are defined below.

The term "naturally-occurring" as used herein as applied to an object refers to the fact that an object can be found in nature. For example, a polypeptide or polynucleotide sequence that is present in an organism (including viruses) that can be isolated from a source in nature and which has not been intentionally modified by man in the laboratory is naturally-occurring. As used herein, laboratory strains of rodents which may have been selectively bred according to classical genetics are considered naturally-occurring animals.

The term "glycoconjugate DNA ligand" as used herein refers to DNA-binding compounds comprising a structural backbone of calicheamicin $\gamma 1^I$ or esperamicin A_1 and like compounds and lacking a reactive enediyne core substituent. Preferably, a glycoconjugate DNA ligand comprises a species of Structure I and binds DNA in a sequence-selective manner; calicheamicin MG is a preferred species of glycoconjugate DNA ligand for some embodiments (e.g., NFAT recognition site binding).

The term "sequence-selective binding" as used herein refers to the property of a compound, such as a glycoconjugate DNA ligand, to preferentially bind to one or more DNA sequences of more than two and less than 100 nucleotides in length as compared to the compound's binding affinity (e.g., Kd) to a pool of random sequence DNA or to bulk human, calf, or salmon genomic

DNA. For example, a compound such as calicheamicin $\gamma 1^I$ may preferentially bind to the tetranucleotide sequence TCCT, TCCC, TCCA, ACCT, TCCG, GCCT, CTCT, TCTC, or TTTT as compared to the remaining tetranucleotide sequence combinations, thus exhibiting sequence-selective binding. Within a group of sequences that are selectively bound, a subset of those sites may be more preferentially bound (i.e., have a higher binding affinity). Calicheamicin $\gamma 1^I$ preferentially binds to homopyrimidine/homopurine sites, such as 5'-TCCT/AGGA and 5'-TCTC/GAGA.

The term "transcriptional enhancement" is used herein to refer to functional property of producing an increase in the rate of transcription of linked sequences that contain a functional promoter.

The term "agent" is used herein to denote a chemical compound, a mixture of chemical compounds, a biological macromolecule, or an extract made from biological materials such as bacteria, plants, fungi, or animal (particularly mammalian) cells or tissues. Agents are evaluated for potential activity as transcriptional modifiers (e.g., immunosuppressants, antineoplastics, etc.) by inclusion in screening assays described hereinbelow. Agents may be selected for their ability to displace a sequence-selective glycoconjugate DNA ligand from a predetermined polynucleotide sequence; such agents may be administered in conjunction with a sequence-selective glycoconjugate DNA ligand for various uses, such as an antidote to reverse the effects of a previously administered glycoconjugate DNA ligand, or as a sharpening agent to enhance the binding preference of the glycoconjugate DNA ligand for a particular predetermined DNA sequence as compared to other similar sequences.

The terms "immunosuppressant" and "immunosuppressant agent" are used herein interchangeably to refer to agents that have the functional property of inhibiting an immune response in human, particularly an immune response that is mediated by activated T-cells.

As used herein, a "transcription factor recognition

site" and a "transcription factor binding site" refer to a polynucleotide sequence(s) or sequence motif(s) which are identified as being sites for the sequence-specific interaction of one or more transcription factors, frequently taking the form of direct protein-DNA binding. Typically, transcription factor binding sites can be identified by DNA footprinting, gel mobility shift assays, and the like, and/or can be predicted on the basis of known consensus sequence motifs, or by other methods known to those of skill in the art. For example and not to limit the invention, eukaryotic transcription factors include, but are not limited to: NFAT, AP1, AP-2, Sp1, OCT-1, OCT-2, OAP, NFκB, CREB, CTF, TFIIA, TFIIB, TFIID, Pit-1, C/EBP, SRF (Mitchell PJ and Tijan R (1989) Science 245: 371). For purposes of the invention, steroid receptors, RNA polymerases, and other proteins that interact with DNA in a sequence-specific manner and exert transcriptional regulatory effects are considered transcription factors.

The term "NFAT-dependent gene" is used herein to refer to genes which: (1) have a NFAT recognition site within 10 kilobases of the first coding exon of said gene, and (2) manifest an altered rate of transcription, either increased or decreased, from a major or minor transcriptional start site for said gene, wherein such alteration in transcriptional rate correlates with the presence of NFAT.

The terms "transcriptional modulation" and "transcriptional influence" are used herein to refer to the capacity to either enhance transcription or inhibit transcription of a structural sequence linked in cis; such enhancement or inhibition may be contingent on the occurrence of a specific event, such as stimulation with an inducer and/or may only be manifest in certain cell types. For example but not for limitation, expression of a protein that prevents formation of an activated (e.g., ligand-bound) glucocorticoid receptor will alter the ability of a glucocorticoid-responsive cell type to modulate transcription of an glucocorticoid-responsive gene in the presence of glucocorticoid. This alteration will be manifest as an inhibition of the transcriptional enhancement of the

glucocorticoid-responsive gene that normally ensues following stimulation with glucocorticoids. The altered ability to modulate transcriptional enhancement or inhibition may affect the inducible transcription of a gene, such as in the just-cited example, or may effect the basal level transcription of a gene, or both. For example, a reporter polynucleotide may comprise a glucocorticoid-inducible enhancer-promoter directing transcription of a sequence encoding a reporter protein (e.g., luciferase, β-galactosidase, chloramphenicol acetyltransferase). Such a reporter polynucleotide may be transferred to a glucocorticoid-responsive cell line for use as a reporter host cell to screen a panel of glycoconjugate DNA ligands for the ability to affect a glucocorticoid-induced transcriptional modulation by sequence-selective interaction with one or more glucocorticoid DNA recognition site sequences. Glycoconjugate DNA ligands that enhance transcription of the cis-linked reporter gene in the absence of glucocorticoids may be identified as putative positive regulators of glucocorticoid-response elements, whereas glycoconjugate DNA ligands that silence expression of the reporter in cells cultured in the presence of glucocorticoids may be identified (e.g., by selecting cells not expressing the cell surface reporter) as glycoconjugate DNA ligands that interfere with the expression of glucocorticoid-responsive genes. Numerous other specific examples of transcription regulatory elements, such as specific enhancers and silencers, are known to those of skill in the art and may be selected for use in the methods and polynucleotide constructs of the invention on the basis of the practitioner's desired application. Literature sources and published patent documents, as well as GenBank™ and other sequence information data sources can be consulted by those of skill in the art in selecting suitable transcription regulatory elements for use in the invention. Where necessary, a transcription regulatory element may be constructed by synthesis (and ligation, if necessary) of oligonucleotides made on the basis of available sequence information (e.g., GenBank sequences for a CD4 enhancer or a SV40 early promoter). For example but not for limitation, a glycoconjugate DNA ligand that prevents

transcriptional activation of NFAT-dependent genes in the presence of functional NFAT generally will alter the ability of a T cell to modulate transcription of an IL-2 gene in response to an antigen stimulus. This alteration will be manifest as an inhibition of the transcriptional enhancement of the IL-2 gene that normally ensues following T cell stimulation. The altered ability to modulate transcriptional enhancement or inhibition may affect the inducible transcription of a gene, such as in the just-cited IL-2 example, or may effect the basal level transcription of a gene, or both. Agents that disrupt, for example, binding of silencer proteins to silencer transcription regulatory elements typically produce an increase in basal and/or induced transcription rate of a cis-linked gene.

As used herein, the term "operably linked" refers to a linkage of polynucleotide elements in a functional relationship. A nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For instance, a promoter or enhancer is operably linked to a coding sequence if it affects the transcription of the coding sequence. Operably linked means that the DNA sequences being linked are typically contiguous and, where necessary to join two protein coding regions, contiguous and in reading frame. However, since enhancers generally function when separated from the promoter by several kilobases and intronic sequences may be of variable lengths, some polynucleotide elements may be operably linked but not contiguous.

As used herein, the terms "expression cassette" and "reporter polynucleotide" refer to a polynucleotide comprising a promoter sequence and, optionally, an enhancer and/or silencer element(s), operably linked to a structural sequence, such as a cDNA sequence or genomic DNA sequence encoding a reporter protein (e.g., luciferase, β-galactosidase, chloramphenicol acetyltransferase), such that the reporter gene sequence is under the transcriptional influence of a cis-acting transcription factor binding site and/or recognition site. In some embodiments, an expression cassette may also include polyadenylation site sequences to ensure polyadenylation of

transcripts. When an expression cassette is transferred into a suitable host cell, the structural sequence is transcribed from the expression cassette promoter, and a translatable message is generated, either directly or following appropriate RNA splicing.

As used herein, the term "reporter host cell" refers to a eukaryotic cell, preferably a mammalian cell, which harbors a reporter polynucleotide. Preferably, the reporter polynucleotide is stably integrated into a host cell chromosomal location, either by nonhomologous integration or by homologous sequence targeting, although transient transfection methods may be employed.

As used herein, the term "endogenous DNA sequence" refers to naturally-occurring polynucleotide sequences contained in a eukaryotic cell. Such sequences include, for example, chromosomal sequences (e.g., structural genes, promoters, enhancers, recombinatorial hotspots, repeat sequences, integrated proviral sequences). A "predetermined sequence" is a sequence which may be selected at the discretion of the practitioner on the basis of known or predicted sequence information. An exogenous polynucleotide is a polynucleotide which is transferred into a eukaryotic cell.

The term "alkyl" refers to a cyclic, branched, or straight chain alkyl group containing only carbon and hydrogen, and unless otherwise mentioned, contain one to twelve carbon atoms. This term is further exemplified by groups such as methyl, ethyl, n-propyl, isobutyl, t-butyl, pentyl, pivalyl, heptyl, adamantyl, and cyclopentyl. Alkyl groups can either be unsubstituted or substituted with one or more substituents, e.g., halogen, alkyl, alkoxy, alkylthio, trifluoromethyl, acyloxy, hydroxy, mercapto, carboxy, aryloxy, aryl, arylalkyl, heteroaryl, amino, alkylamino, dialkylamino, morpholino, piperidino, pyrrolidin-1-yl, piperazin-1-yl, or other functionality.

The term "lower alkyl" refers to a cyclic, branched or straight chain monovalent alkyl radical of one to six carbon atoms. This term is further exemplified by such radicals as methyl, ethyl, n-propyl, i-propyl, n-butyl, t-butyl, i-butyl (or 2-methylpropyl), cyclopropylmethyl, i-amyl, n-amyl, and hexyl.

The term "aryl" or "Ar" refers to a monovalent unsaturated aromatic carbocyclic group having a single ring (e.g., phenyl) or multiple condensed rings (e.g., naphthyl or anthryl), which can optionally be unsubstituted or substituted with, e.g., halogen, alkyl, alkoxy, alkylthio, trifluoromethyl, acyloxy, hydroxy, mercapto, carboxy, aryloxy, aryl, arylalkyl, heteroaryl, amino, alkylamino, dialkylamino, morpholino, piperidino, pyrrolidin-1-yl, piperazin-1-yl, or other functionality.

The term "substituted alkoxy" refers to a group having the structure -O-R, where R is alkyl which is substituted with a non-interfering substituent. The term "arylalkoxy" refers to a group having the structure -O-R-Ar, where R is alkyl and Ar is an aromatic substituent. Arylalkoxys are a subset of substituted alkoxys. Examples of preferred substituted alkoxy groups are: benzyloxy, napthyloxy, and chlorobenzyloxy.

The term "aryloxy" refers to a group having the structure -O-Ar, where Ar is an aromatic group. A preferred aryloxy group is phenoxy.

The term "non-interfering substituent" as used herein refers to a chemical substituent which does not comprise an enediyne structure or core aglycone (i.e., calicheamicinone substituent) and which does not, when present in a compound, result in the compound exhibiting a substantially reduced (e.g., by more than two orders of magnitude) affinity for binding DNA in a sequence-selective manner. Frequently, non-interfering substituents can comprise moieties which enhance pharmacological or pharmacokinetic properties (e.g., hydrophobic moieties such as aryl or naphthalene groups, and the like).

DETAILED DESCRIPTION

A basis of the invention is the unexpected finding that structural variants of naturally occurring eyediyne anticancer antibiotics can interact with DNA in a sequence-selective manner, and this sequence-selective interaction can result in the displacement or inhibition of binding of certain DNA-binding proteins (e.g., transcription factors) to their recognition site

sequences. Glycoconjugate DNA ligands that comprise the structural backbone (i.e., the saccharide/aryl backbone) of calicheamicin $\gamma 1^I$ and other naturally occurring enediyne antibiotics (e.g., esperimicin A_1), and which lack the reactive enediyne moiety of the naturally occurring antibiotic, can possess sequence-selective DNA binding activity and lack DNA cleavage activity. These non-reactive glycoconjugate DNA ligands can possess sequence-selective, reversible noncovalent binding to DNA, and may be used to alter the interactions of certain DNA-binding proteins (e.g., transcription factors) to their DNA recognition sequences on a sequence-selective basis. The targeted alteration of protein-DNA binding interactions by the glycoconjugate DNA ligands of the invention serve as the basis for selective elution of transcription factors from predetermined DNA binding sites, for transcriptional modulation of genes under the transcriptional influence of a DNA sequence to which the glycoconjugate DNA ligand preferentially binds, and for other embodiments.

A preferred embodiment is the calicheamicin-based glycoconjugate DNA ligand, calicheamicin MG shown in Fig. 1, which binds in a sequence-selective manner to the tetranucleotide sequences TCCT, TCCC, TCCA, ACCT, TCCG, GCCT, CTCT, TCTC, or TTTT, especially to 5'-TCCT-3' (and its complement 5'-AGGA-3') and 5'-TCTC-3' (and its complement 5'-GAGA-3'), and which preferentially binds to NFAT recognition site sequences interfering with the interaction of NFAT to its recognition site(s), thereby modulating NFAT-dependent transcription.

DNA Binding Proteins

DNA-binding proteins perform fundamental functions in cell biology and in organismal physiology. DNA-binding proteins are involved in a variety of activities, including replicating the genome, transcribing active genes, repressing inactive genes, repairing damaged DNA, and as basic components of chromatin structure. In particular, transcription factors regulate cell development, differentiation, and cell growth by binding to a specific DNA site (or set of sites) and regulating gene transcription. Critical to the function of transcription factors

is their ability to selectively bind to specific DNA sequence motifs located in or near structural gene sequences and regulatory elements (e.g., enhancers, silencers, promoters).

Many DNA-binding proteins can be grouped into classes that use related structural motifs for recognition. Large, well-established families include, for example, the helix-turn-helix (HTH) proteins, the homeodomains, zinc finger proteins, steroid receptor DNA-binding domains, leucine zipper proteins, and the helix-loop-helix proteins, and β-sheet DNA-binding proteins (Pabo CO and Sauer RT (1992) Ann. Rev. Biochem. 61: 1053). There are many important DNA-binding proteins that are not categorized into the known structural families; the SV40 large T antigen and the human p53 tumor suppressor gene are two such examples.

Transcription factors in particular tend to possess sequence-specific binding properties which are associated with their activities as transcriptional regulators of adjacent structural genes. The precise physical basis of sequence-specific binding is not fully understood with regard to all known transcription factors, and the nature of sequence-selectivity may be very complex and require considerations of local variations in DNA structure and/or flexibility. However, most of the well-characterized families of DNA-binding proteins use α-helices to make base contacts in the major groove of B-form DNA, although β-sheets or regions of extended polypeptide chain can also be used to make base contacts.

Hydrogen bonding between peptide side chains and nucleic acid bases are frequently involved in sequence-specific recognition and binding, although contacts with the DNA backbone, especially hydrogen bonds to phosphodiester oxygens and hydrophobic interactions with the sugar rings in the DNA backbone, are often present as well. Structural studies of contacts with bases have revealed (1) direct hydrogen bonds between the protein side chains and the bases, (2) hydrogen bonds between the polypeptide backbone and the bases, (3) hydrogen bonds mediated by water molecules, and (4) hydrophobic contacts (Pabo CO and Sauer RT (1992) op.cit.). Most base contacts appear to occur through the major groove of B-form DNA, although there

could be sequence-dependent influences on the local structure of a given binding site that deviates from canonical B-form DNA. Furthermore, there does not appear to be a simple "recognition code" wherein a one-to-one correspondence between the amino acid side chains and the bases or DNA backbone moieties they contact, further implicating the importance of local structure as a basis for sequence-specific recognition.

NFAT-Dependent Transcriptional Activity

One class of transcription factor binding sites are the DNA recognition sequences for the NFAT transcription factor, which is involved in T cell activation and the like.

The immune response is coordinated by the actions of cytokines produced from activated T lymphocytes. The precursors for most T lymphocytes arise in the bone marrow and migrate to the thymus where they differentiate and express receptors capable of interacting with antigen. These differentiated T lymphocytes then migrate to the peripheral lymphoid organs where they remain quiescent until they come in contact with the cognate antigen. The interaction of antigen with the antigen receptor on T lymphocytes initiates an ordered series of pleiotropic changes; a process denoted as T lymphocyte activation. T lymphocyte activation is a 7 to 10 day process that results in cell division and the acquisition of immunological functions such as cytotoxicity and the production of lymphokines that induce antibody production by B lymphocytes and control the growth and differentiation of granulocyte and macrophage precursors. The cytokines produced by activated T lymphocytes act upon other cells of the immune system to coordinate their behavior and bring about an effective immune response.

The initiation of T lymphocyte activation requires a complex interaction of the antigen receptor with the combination of antigen and self-histocompatibility molecules on the surface of antigen-presenting cells.

T lymphocyte activation involves the specific regulation of particular subsets of genes. The transcriptional regulation characteristic of T cell activation begins minutes

after the antigen encounter and continues until at least 10 days later. The T lymphocyte activation genes can be grouped according to the time after stimulation at which each gene is transcribed. Early genes are the first subset of T lymphocyte activation genes that is expressed during the activation process. Expression of the early genes triggers the transcriptional modulation of subsequent genes in the activation pathway. Because of the critical role of the T lymphocyte in the immune response, agents that interfere with expression of the early activation genes, such as cyclosporin A and FK506, are effective immunosuppressants.

The induction of cytokine production in T lymphocytes as a consequence of specific contact with antigens serves to coordinate the immune response. Cytokines are responsible for the control of proliferation and differentiation among precursors of B cells, granulocytes, and macrophages. One such cytokine that is produced by activated T cells is interleukin-2 (IL-2).

The IL-2 gene is essential for both the proliferation and immunologic function of T cells. Moreover, the IL-2 gene is a representative and important early gene that is transcribed at an enhanced rate during T cell activation. Putative signalling pathways connect the transcriptional regulation of the IL-2 gene with the antigen receptor on the T cell surface.

Transcription of the early genes requires the presence of specific transcription factors, such as NFAT, which in turn are regulated through interactions with the antigen receptor. These transcription factors are proteins which act through enhancer and promoter elements near the early activation genes to modulate the rate of transcription of these genes. Many of these transcription factors reversibly bind to specific DNA sequences located in and near enhancer elements.

The interleukin-2 (IL-2) gene is a paradigmatic early activation gene. The IL-2 gene product plays a critical role in T lymphocyte proliferation and differentiation. The IL-2 gene is transcriptionally active only in T cells that have been stimulated through the antigen receptor or its associated molecules (Cantrell and Smith (1984) Science 224: 1312). The

transcriptional induction of IL-2 in activated T lymphocytes is mediated by a typical early gene transcriptional enhancer that extends from 325 basepairs upstream of the transcriptional start site for the IL-2 gene (Durand et al. (1988) Mol. and Cell. Biol. 8: 1715). Other genes known to contain NFAT recognition sites in their regulatory regions include: γ-interferon, IL-4, GM-CSF, and others. This region, which is referred to herein as the IL-2 enhancer, has been used extensively to dissect the requirements for T lymphocyte activation. An array of transcription factors, including NFAT, NFkb, AP1, Oct-1, and a newly identified protein that associates with Oct-1 called OAP-40, bind to sequences in this region (Ullman et al. (1991) Science 254: 558-562). These different transcription factors act together to integrate the complex requirements for T lymphocyte activation. One of the functional sequences in the IL-2 enhancer is a binding site for a protein complex, designated NFAT (nuclear factor of activated T lymphocytes), that functions as a transcriptional regulator of IL-2, IL-4, and other early activation genes (Shaw et al. (1988) Science 241: 202-205). The NFAT complex is formed when a signal from the antigen receptor is transduced to the nucleus and an activated NFAT complex forms. Enhancement of transcription of genes adjacent to the NFAT recognition site requires that the NFAT complex bind to the recognition site (Shaw et al. (1988) Science 241: 202-205).

Among the group of transcription factors mentioned above, the presence of NFAT is characteristic of the transcription events involving early activation genes, in that its recognition sequence is able to enhance transcription of linked heterologous genes in activated T cells of transgenic animals (Verweij et al. (1990) J. Biol. Chem. 265: 15788). The NFAT sequence element is also the only known transcriptional element in the IL-2 enhancer that has no stimulatory effect on transcription in the absence of physiologic activation of the T lymphocyte through the antigen receptor or through treatment of T cells with the combination of ionomycin and PMA. For example, the NFAT element enhances transcription of linked sequences in T lymphocytes which have had proper presentation of specific

antigen by MHC-matched antigen presenting cells or have been stimulated with the combination of ionomycin/PMA, but not in unstimulated T lymphocytes (Durand et al., Mol. and Cell. Biol. 8:1715 (1988); Shaw et al., Science 241:202-205 (1988); Karttunen and Shastri, Proc. Natl. Acad. Sci. USA 88:3972-3976 (1991), 1991; Verweij et al., J. Biol. Chem. 265:15788-15795 (1990)). Moreover, the NFAT sequence element naturally enhances transcription of the IL-2 gene only in activated T lymphocytes.

Transcriptional enhancement involving NFAT recognition sequences is completely blocked in T cells treated with efficacious concentrations of cyclosporin A or FK506, with little or no specific effect on transcriptional enhancement involving recognition sites for other transcription factors, such as AP1 and NF-κB (Shaw et al. (1988) Science 241: 202; Emmel et al. (1989) Science 246: 1617; Mattila et al. (1990) EMBO J 9: 4425).

The presence of DNA recognition sequence(s) for the NFAT complex appear to restrict transcription of linked DNA sequences, such as the IL-2 gene and other early activation genes, so that transcription of these linked sequences is enhanced only in stimulated T lymphocytes (Durand et al. (1988) Mol. and Cell. Biol. 8: 1715; Shaw et al. (1988) Science 241: 202; and Verwiej et al. (1990) J. Biol. Chem. 265: 15788). Within the IL-2 enhancer, there are two NFAT recognition sites, a proximal and distal NFAT site. Elimination of an NFAT site from the IL-2 enhancer drastically reduces the ability of the IL-2 enhancer element to function. In addition, tandem multimeric arrays of binding sites for the NFAT protein will direct transcription of linked sequences in activated T lymphocytes or in vitro transcription cocktails containing NFAT heterodimer, but not in other cell types or in transcription cocktails lacking NFAT.

A distinguishing feature of the NFAT DNA binding site is its purine-rich binding site, e.g., 5'-AAGAGGAAAAA-3'. DNA sequence comparisons of the promoter/enhancer regions of several genes that respond to T-cell activation signals has identified putative NFAT protein binding sites. Such a comparison suggests that NFAT or a related family member may bind within the

promoter/enhancer regions of other T-cell activation dependent genes. Most of these genes are sensitive to immunosuppressants, such as FK506 and cyclosporin. A list of putative NFAT binding sites follows in Table I:

TABLE I

Purine Rich Core Sequences	Position	Gene
GAAAGGAGGAAAAACTGTTT	(-289 to -270)	human IL-2
CCAAAGAGGAAAATTTGTTT	(-293 to -274)	murine IL-2
CAGAAGAGGAAAAATGAAGG	(-143 to -124)	human IL-2
TCCAGGAGAAAAAAATGCCTC	(-143 to -124)	human IL-4
AAAACTTGTGAAAATACGTA	(-71 to -52)	human g-IFN
TAAAGGAGAGAACACCAGCT	(-270 to -251)	HIV-LTR
GCAGGGTGGGAAAGGCCTTT	(-241 to -222)	murine GM-CSF

(Abbreviations: IL-2, interleukin 2; IL-4, interleukin 4; HIV-LTR, human immunodeficiency virus long terminal repeat; GM-CSF, granulocyte-macrophage colony stimulating factor.)

The intragenic enhancer of GM-CSF also comprises an NFAT site (Cocherill et al. (1992) Proc. Natl. Acad. Sci. (U.S.A.).

Inhibition of NFAT-dependent transcription can inhibit the process of T cell activation, leading to a reduction in T cell-mediated immune responses such as graft rejection. Unfortunately, cyclosporin A, FK506, and other currently used immunosuppressants possess several disadvantageous properties and produce undesirable side effects. It is believed that direct inhibition of NFAT-dependent gene transcription would provide superior immunosuppressant action and minimize undesirable side-effects. Unfortunately, there is at present no convenient method for selectively inhibiting NFAT-dependent gene transcription that is both efficacious and has minimal side effects.

Glycoconjugate DNA Ligands

One class of small molecule drugs that has been used for human antineoplastic therapy are the naturally-occurring enediyne anticancer antibiotics which bind to and cleave DNA (Nicolaou KC and Dai WM (1991) Angew. Chem. Int. Ed. Engl. 30: 1387). The enediyne family of anticancer antibiotics comprises the neocarzinostatin chromophore (Edo et al. (1985) Tetrahedron

Lett. 26: 331), the calicheamicins (Lee et al. (1987) J. Am. Chem. Soc. 109: 3464), the esperamicins (Golik et al. (1987) J. Am. Chem. Soc. 109: 3461), kedarcidin (Proc. Natl. Acad. Sci. (U.S.A.) (1993) 90: 2822), and the dynemicins (Konishi et al. (1990) J. Am. Chem. Soc. 112: 3715). The DNA binding properties of these enediyne antibiotics have been investigated (Zimmer C and Wahnert U (1986) Prog. Biophys. Molec. Biol. 47: 31; Jarman M (1991) Nature 349: 566; Lu et al. (1991) J. Biomolec. Struct. Dynam. 9: 271; Sugiura et al. (1991) Biochemistry 30: 2989; Walker et al. (1992) Proc. Natl. Acad. Sci. (U.S.A.) 89: 4608; Hawley et al. (1989) Proc. Natl. Acad. Sci. (U.S.A.) 86: 1105; Sugiura et al. (1990) Proc. Natl. Acad. Sci. (U.S.A.) 87: 3831; Dedon PC and Goldberg IH (1992) Biochemistry 31: 1909; Nicolaou et al. (1992) Science 256: 1172; Drak et al. (1991) Proc. Natl. Acad. Sci. (U.S.A.) 88: 7464). Although some of these antibiotics (e.g., calicheamicin, esperimicin) bind DNA in a sequence-selective manner (Drak et al. (1991) op.cit.), they have not been shown to act as transcription modulators for specific genes, nor would they be expected to in view of their propensity to cleave DNA and lead to cytotoxicity.

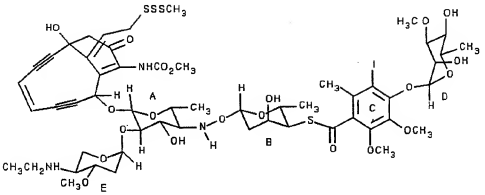
Various enediyne anticancer antibiotics such as calicheamicin $\gamma 1^{I}$, dynemicin A, dynemicin H, calicheamicin T, esperimicin A_1, esperimicin D, and the like can be obtained by standard fermentation and purification methods and/or synthesized or modified by standard organic synthetic procedures (Konishi et al. (1990) J. Am. Chem. Soc. 112: 3715; Konishi et al. (1989) J. Antibiot. 42: 1449; Sugiura et al. (1991) Biochemistry 30: 2989; Lee et al. (1987) J. Am. Chem. Soc. 109: 3464; Lee et al. (1987) J. Am. Chem. Soc. 109: 3466; Golik et al. (1987) J. Am. Chem. Soc. 109: 3462; Haseltine et al. (1991) J. Am. Chem. Soc. 113: 3850; Nicolaou et al. J. Am. Chem. Soc. 112: 8193; Lee et al. (1992) J. Am. Chem. Soc. 114: 985; Walker et al. (1992) Proc. Natl. Acad. Sci. (U.S.A.) 89: 4608; Halcomb et al. (1992) Ang. Chem., Int. Ed. Engl. 31: 338; Halcomb et al. (1992) Ang. Chem. 104: 314; Halcomb et al. (1991) J. Am. Chem. Soc. 113: 5080; Nicolaou et al. (1992) Ang. Chem., Int. Ed. Engl. 31: 855; Nicolaou et al. Ang. Chem. 104: 926; and Smith et al. (1992) J.

Amer. Chem. Soc. 114: 3134.

The calicheamicins and esperimicins preferentially recognize specific sequences of nucleotides in double-stranded DNA, and cleave the chains at preferred sites. The structural basis for the sequence selectivity of the calicheamicins and esperimicins is not known, however interactions between the sugar portions of these molecules and specific sites on DNA may be involved (Zein et al. (1989) Science 244: 697; Walker et al. (1992) Proc. Natl. Acad. Sci. (U.S.A.) 89: 4608; Drak et al. (1991) Proc. Natl. Acad. Sci. (U.S.A.) 88: 7464). Calicheamicin $\gamma 1^{I}$ binds to DNA in a sequence-selective manner and then produces strand scission via rearrangement of its enediyne core to form a reactive biradical.

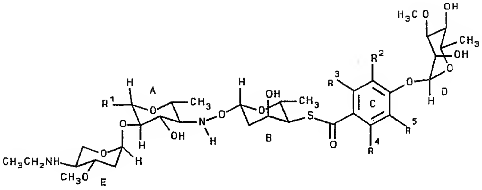
Enediyne compounds potently cleave DNA through bioreduction of the trisulfide and subsequent 1,4-benzenoid diradical formation, which results in hydrogen abstraction from adjacent nucleotide bases. The subsequent double strand DNA cleavage occurs at sites on each strand that are separated by three base pairs, suggesting a minor groove interaction. Modeling studies indicate that the stereochemical orientation of the aglycone relative to the carbohydrate recognition element permits a complementarity of fit on placement of calicheamicin into the minor groove and that specific interactions between the iodo substituent on the aromatic ring and adjacent N2 amino substituents of the guanine dinucleotide may also contribute to the preferred sequence specificity.

The structure of calicheamicin $\gamma 1^{I}$ is shown below, with the calicheamicin polysaccharide/aryl backbone composed of sugars designated herein as rings A, B, D, and E and a thiobenzoate moiety designated as ring C. The enediyne core moiety (a cyclic enediyne) is linked to the A sugar ring through an ether linkage. The eyediyne is capable of rearranging to generate an aromatic ring structure with two chemically active radical sites (a 1,4-benzenoid diradical) following bioreduction of the trisulfide residue of the enediyne core. The two unbonded electrons of the biradical ring can abstract hydrogen atoms from the sugar phosphate backbone of DNA, resulting in double strand breakage.



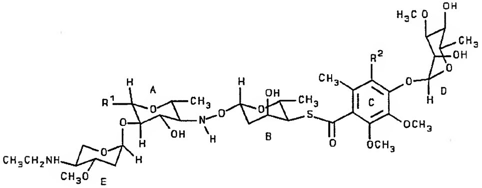
In one aspect of the invention, glycoconjugate DNA ligands having the backbone structure of calicheamicin (i.e., comprising rings A, B, C, D, and/or E) and lacking the reactive enediyne core moiety are provided for modulating transcription of a gene by sequence-selective DNA binding to a DNA-binding protein recognition sequence. Such binding of the glycoconjugate DNA ligand to the recognition site(s) thereby alters (typically by inhibiting) binding of a DNA-binding protein to its recognition site and affects (usually negatively) transcription dependent upon the DNA-binding protein (e.g., NFAT-dependent transcription if the glycoconjugate binds to a NFAT site). In alternative embodiments, the glycoconjugate DNA ligands comprise a subset of the backbone structure of calicheamicin, such as comprising rings A, B, and C but lacking the D and E rings; or comprising rings E, A, and B but lacking the C and D rings, as well as other variations. Glycoconjugate DNA ligands having a variety of backbone structures other than calicheamicin and esperimicin are also within the scope of the present invention, although glycoconjugate DNA ligands comprising a backbone structure of calicheamicin are believed preferred for inhibition of NFAT-dependent transcription and sequence-selective binding to NFAT recognition sites.

A preferred class of such calicheamicin-related glycoconjugate DNA ligands have the general structural formula (Structure I):



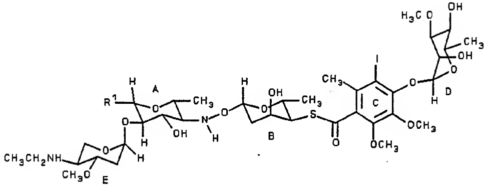
where R_1 is hydrogen, halogen, hydroxy, lower alkyl, lower alkoxy, or aryloxy, preferably methyl, ethyl, propyl, butyl, methoxy, ethoxy, propyloxy, phenoxy, or benzyloxy; R_2 is a halogen, hydroxy, or lower alkyl, preferably iodo, bromo, chloro, fluoro, methyl, or ethyl; R_3 is lower alkyl or lower alkoxy, preferably methyl, ethyl, methyloxy, or ethyloxy, or hydrogen; and R_4 and R_5 are independently selected from lower alkyl or lower alkoxy, preferably methyl, ethyl, methyloxy, ethyloxy, or hydrogen. It is believed that, in alternative embodiments, the R_1 substituent may comprise any of a variety of non-interfering substituents (e.g., alkyl, aryl, arylalkyl, aryloxy, substituted alkoxy, alkanol, halogen, heterocyles, or hydrogen). The ethylamino group in Structure I and derived subclasses can be optionally replaced by R_6-N-R_7, wherein R_6 and R_7 are hydrogen, lower alkyl, or alkyl.

A particularly preferred class of calicheamicin-related glycoconjugate DNA ligands comprise the following structural formula (Structure II):



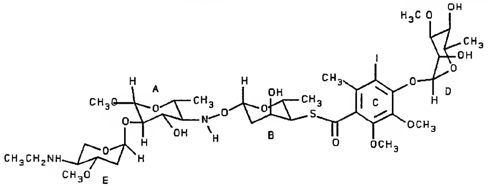
where R_1 is a non-interfering substituent, preferably hydrogen, hydroxy, lower alkyl, lower alkoxy, aryloxy, or alkanol, more preferably methyl, ethyl, propyl, butyl, methoxy, ethoxy, propyloxy, or ethylhydroxyl; and R_2 is a halogen, hydrogen, lower alkyl or lower alkoxy, preferably iodo, bromo, fluoro, chloro, methyl, ethyl, or methoxy. Typically, R_1 is hydroxy, methoxy, or ethoxy. Typically, R_2 is iodo, bromo, chloro, or fluoro; most typically iodo.

A preferred subclass of calicheamicin-related glycoconjugate DNA ligands have the structural formula (Structure III):



where R_1 is a non-interfering substituent, preferably hydrogen, hydroxy, lower alkyl, lower alkoxy, aryloxy, or alkanol, more preferably methyl, ethyl, propyl, butyl, methoxy, ethoxy, propyloxy, hydroxy, halogen, or hydrogen.

An especially preferred species of Structure I, Structure II, and Structure III is the methyl glycoside derivative of the aryltetrasaccharide, calicheamicin MG (CLM-MG):



Calicheamicin MG exhibits sequence-selective binding to DNA (e.g., preferentially binding to TCCT, TTTT) and substantially lacks DNA cleavage activity. Calicheamicin MG lacks the enediyne core moiety present in calicheamicin $\gamma 1^{I}$ that forms a reactive biradical species upon bioreduction (e.g., with NADPH); thus, calicheamicin MG substantially lacks DNA cleavage activity but exhibits reversible non-covalent binding to DNA in a sequence-selective manner. Calicheamicin MG and other sequence-selective congeners according to Structures I, II, and III can be used to displace, disrupt, or prevent binding of a DNA-binding protein, such as a sequence-specific transcription factor (e.g., NFAT), to their cognate DNA recognition sites which comprise a sequence (e.g., TCCT) to which calicheamicin MG (or other calicheamicin-related glycoconjugate DNA ligand of Structure I, II, or III) preferentially binds. These calicheamicin-related glycoconjugate DNA ligands do not contain reactive enediyne moieties and are thus suited for various uses, including as pharmaceuticals and commercial research reagents.

The sequence-selectivity of the calicheamicin-related glycoconjugate DNA ligands of the invention can vary on the basis of the primary structure of the glycoconjugate DNA ligand. This provides an advantageous basis for choosing non-reactive DNA-binding agents which possess reversible sequence-selective binding to a particular predetermined DNA sequence. For example, calicheamicin MG preferentially to the TCCT and TTTT sequences which are present in, for example, an NFAT recognition site. Calicheamicin MG acts as a reversible ligand for binding to the TTTT and/or TCCT sequence(s) in an NFAT recognition site and inhibits binding of an NFAT complex to its recognition site. This inhibition of binding of NFAT to its recognition site modulates transcription, typically by inhibiting NFAT-dependent transcription of NFAT-dependent structural genes (e.g., the human IL-2 gene). This advantageous feature of calicheamicin MG provides the basis for its use as a modulator of the transcription of NFAT-dependent genes both in vivo and in vitro, and for the use of calicheamicin MG as a commercial research reagent for the elution of preparative quantities of NFAT

protein(s) from DNA affinity chromatography resin employing NFAT recognition sites in the DNA, among other uses.

Sequence-Selective Glycoconjugate DNA Ligands

Structural congeners of esperimicin and calicheamicin which lack reactive enediyne moieties can by synthesized according to conventional synthetic methods known to those skilled in the art, as well as other methods provided herein. Preferred structural congeners of calicheamicin are those having a structural formula of Structure I, II, or III. Other structural congeners comprise, for example, the basic structure of calicheamicin MG or Structure I, II, or III but which lack: (1) the E sugar ring, (2) the D sugar ring, or (3) both the D and E sugar rings. Additional structural congeners comprising the basic structure of CLM-MG or of structures I, II, or III, but which have been multimerized in varying combinations (e.g., strcture I linked to structure II, etc.) may be used.

The degree of sequence-selective binding exhibited by each such structural congener is determined by assay, wherein a predetermined DNA-binding protein (preferably a transcription factor) and at least one polynucleotide comprising a known recognition sequence for the predetermined DNA-binding protein are contacted under binding conditions in the presence of the structural congener. Typically, the binding reaction also contains at least one alternative species of DNA-binding protein and at least one alternative polynucleotide having a cognate recognition site for the alternative DNA-binding protein; the predetermined DNA-binding protein and the alternative DNA-binding protein(s) bind to distinct DNA sequences. Optionally, the alternative DNA-binding protein(s) and the cognate alternative DNA recognition site(s) may comprise a separate, parallel binding reaction. A control set of binding reactions lacking the structural congener are included for calibration of binding conditions.

Usually, the polynucleotide comprising the recognition site(s) for the predetermined DNA-binding protein is distinguishable from the polynucleotide comprising the

alternative DNA-binding protein recognition site(s) (e.g., they may be of different lengths and electrophoretically separable). Usually, one or more of the polynucleotides is/are labelled with a detectable label (e.g., ^{32}P, ^{35}S) and/or one or more of the DNA-binding proteins (the predetermined DNA-binding protein and/or an alternative DNA-binding protein) is/are also labeled with a detectable label (e.g., radioiodination with ^{125}I or ^{131}I by chloramine T/Bolton-Hunter reaction, radiolabeling by incorporation of ^{35}S-methionine, ^{14}C-leucine, and the like). The differential ability of the structural congener to inhibit or displace binding of the predetermined DNA-binding protein to its cognate recognition sequence as compared to the structural congener's ability to inhibit or displace binding of the alternative DNA-binding protein(s) to its/their cognate DNA recognition site(s) is determined. Various methods for performing such a determination of differential binding inhibition are available.

Binding assays generally take one of two forms: immobilized target DNA having the recognition sequence can be used to bind labeled DNA-binding protein(s), or conversely, immobilized DNA-binding protein(s) can be used to bind labeled target DNA comprising the cognate recognition sequence(s). In each case, the labeled macromolecule (protein or DNA) is contacted with the immobilized macromolecule (respectively, DNA or protein) under aqueous conditions that permit specific binding of the DNA-binding protein(s) to the target DNA in the absence of the structural congener (glycoconjugate DNA ligand). Particular aqueous conditions may be selected by the practitioner according to conventional methods, including methods employed in DNA-protein footprinting and/or in vitro nuclear run-on transcription. For general guidance, the following buffered aqueous conditions may be used: 20-200 mM NaCl, 5-50 mM Tris HCl, pH 5-8, and nanomolar to millimolar concentrations of Zn^{+2} and, optionally, Mg^{+2} and/or Mn^{+2}. It is appreciated by those in the art that additions, deletions, modifications (such as pH) and substitutions (such as KCl substituting for NaCl or buffer substitution) may be made to these basic conditions.

Modifications can be made to the basic binding reaction conditions so long as specific binding of the DNA-binding protein(s) to target DNA occurs in the control reaction(s). Conditions that do not permit specific binding in control reactions (no glycoconjugate DNA ligand included) are not suitable for use in DNA binding assays. Conventional electrophoretic "mobility shift" assays (Fried M and Crothers D (1981) Nucleic Acids Res. 9: 6505; Garner M and Rezvin A (1981) Nucleic Acids Res. 9: 3407; Strauss F and Varshavsky A (1984) Cell 37: 889) may be used to calibrate the binding reaction conditions for satisfactory binding of reagents. However, it is believed that assays using an immobilized component (DNA or DNA-binding protein) will be preferred for most large-scale screening applications.

In embodiments where target DNA containing a DNA-binding protein recognition site is immobilized, preferably double-stranded DNA containing at least one recognition site sequence is bonded, either covalently or noncovalently, to a substrate. For example, but not for limitation, DNA can be covalently linked to a diazotized substrate, such as diazotized cellulose, particularly diazophenylthioether cellulose and diazobenzyloxymethyl cellulose (Alwine et al. (1977) Proc. Natl. Acad. Sci. (U.S.A.) 74: 5350; Reiser et al. (1978) Biochem. Biophys. Res. Commun. 85: 1104; Stellwag and Dahlberg (1980) Nucleic Acids Res. 8: 299, which are incorporated herein by reference). Alternatively, DNA can be covalently linked to a substrate by partial ultraviolet light-induced crosslinking to a Nylon 66 or nitrocellulose substrate (Church and Gilbert (1984) Proc. Natl. Acad. Sci. (U.S.A.) 81: 1991, which is incorporated herein by reference). Also, for example and not for limitation, DNA can be noncovalently bound to a Nylon 66 or other highly charged anionic substrate (Berger and Kimmel, Methods in Enzymology, Volume 152, Guide to Molecular Cloning Techniques (1987), Academic Press, Inc., San Diego, CA). In some embodiments, it is preferable to use a linker or spacer to reduce potential steric hindrance from the substrate. The immobilized DNA is contacted with the labeled DNA-binding protein under

aqueous binding conditions.

Preferably, at least one DNA-binding protein species is labeled with a detectable marker. Suitable labeling includes, but is not limited to, radiolabeling by incorporation of a radiolabeled amino acid (e.g., ^{14}C-labeled leucine, ^{3}H-labeled glycine, ^{35}S-labeled methionine), radiolabeling by post-translational radioiodination with ^{125}I or ^{131}I (e.g., Bolton-Hunter reaction and chloramine T), labeling by post-translational phosphorylation with ^{32}P (e.g., phosphorylase and inorganic radiolabeled phosphate) fluorescent labeling by incorporation of a fluorescent label (e.g., fluorescein or rhodamine), or labeling by other conventional methods known in the art. Isotopes of Zn (Mn or Mg) may also be used, such as for zinc-finger proteins, and the like. In embodiments where the target DNA is immobilized by linkage to a substrate, the predetermined DNA-binding protein (and the alternative DNA-binding protein) is labeled with a detectable marker.

Additionally, in some embodiments a DNA-binding protein may be used in combination with an accessory protein (e.g., a protein which forms a transcription complex with the DNA-binding protein in vivo), it is preferred that different labels are used for each polypeptide species, so that binding of individual and/or heterodimeric and/or multimeric complexes to target DNA can be distinguished, as can binding of the alternative DNA-binding protein(s), if present in the same binding reaction. For example but not limitation, a predetermined DNA-binding protein may be labeled with fluorescein and an accessory polypeptide (or an alternative DNA-binding protein) may be labeled with a fluorescent marker that fluorescesces with either a different excitation wavelength or emission wavelength, or both. Alternatively, double-label scintillation counting may be used, wherein the predetermined DNA-binding protein is labeled with one isotope (e.g., ^{3}H) and a second polypeptide species (accessory protein or alternative DNA-binding protein) is labeled with a different isotope (e.g., ^{14}C) that can be distinguished by scintillation counting using discrimination techniques.

Labeled DNA-binding proteins are contacted with

immobilized target DNA under aqueous conditions as described *infra*. The time and temperature of incubation of a binding reaction may be varied, so long as the selected conditions permit specific binding to occur in a control reaction where no agent is present. Preferable embodiments employ a reaction temperature of about at least 15 degrees Centigrade, more preferably 35 to 42 degrees Centigrade, and a time of incubation of approximately at least 15 seconds, although longer incubation periods are preferable so that, in some embodiments, a binding equilibrium is attained. Binding kinetics and the thermodynamic stability of bound protein:DNA complexes determine the latitude available for varying the time, temperature, salt, pH, and other reaction conditions. However, for any particular embodiment, desired binding reaction conditions can be calibrated readily by the practitioner using conventional methods in the art, which may include binding analysis using Scatchard analysis, Hill analysis, and other methods (<u>Proteins, Structures and Molecular Principles</u>, (1984) Creighton (ed.), W.H. Freeman and Company, New York).

Specific binding of labeled DNA-binding protein to immobilized DNA is determined by including unlabeled competitor protein(s) (e.g., albumin) and/or free competitor DNA or competitor oligonucleotides. After a binding reaction is completed, labeled DNA-binding protein(s) that is/are specifically bound to immobilized target DNA is detected. For example and not for limitation, after a suitable incubation period for binding, the aqueous phase containing non-immobilized protein and nucleic acid is removed and the substrate containing the target DNA and any labeled protein bound to the DNA is washed with a suitable buffer, optionally containing unlabeled blocking agent(s), and the wash buffer(s) removed. After washing, the amount of detectable label remaining specifically bound to the immobilized DNA is determined (e.g., by optical, enzymatic, autoradiographic, or other radiochemical methods).

In some embodiments, addition of unlabeled blocking agents that inhibit non-specific binding are included. Examples of such blocking agents include, but are not limited to, the following: calf thymus DNA, salmon sperm DNA, yeast RNA, mixed

sequence (random or pseudorandom sequence) oligonucleotides of various lengths, bovine serum albumin, nonionic detergents (NP-40, Tween, Triton X-100, etc.), nonfat dry milk proteins, Denhardt's reagent, polyvinylpyrrolidone, Ficoll, and other blocking agents. Practioners may, in their discretion, select blocking agents at suitable concentrations to be included in DNA binding assays; however, reaction conditions are selected so as to permit specific binding between the predetermined DNA-binding protein and target DNA containing its recognition sequence in a control binding reaction. Blocking agents are included to inhibit nonspecific binding of labeled DNA-binding protein to immobilized DNA and/or to inhibit nonspecific binding of labeled DNA to immobilized DNA-binding protein.

In embodiments where the DNA-binding protein is immobilized, covalent or noncovalent linkage to a substrate may be used. Covalent linkage chemistries include, but are not limited to, well-characterized methods known in the art (Kadonaga and Tijan (1986) Proc. Natl. Acad. Sci. (U.S.A.) 83: 5889, which is incorporated herein by reference). One example, not for limitation, is covalent linkage to a substrate derivatized with cyanogen bromide (such as CNBr-derivatized Sepharose 4B). It may be desirable to use a spacer to reduce potential steric hindrance from the substrate. Noncovalent bonding of proteins to a substrate include, but are not limited to, bonding of the protein to a charged surface and binding with specific antibodies. DNA is typically labeled by incorporation of a radiolabeled nucleotide (^{3}H, ^{14}C, ^{35}S, ^{32}P) or a biotinylated nucleotide that can be detected by labeled avidin (e.g., avidin containing a fluorescent marker or enzymatic activity).

Frequently, a gel mobility shift assay is performed wherein the bound complex comprising the predetermined DNA-binding protein and the polynucleotide comprising the predetermined recognition site can be distinguished from unbound polynucleotide comprising the predetermined recognition site; also usually the bound complex comprising the alternative DNA-binding protein and the polynucleotide comprising the cognate alternative recognition site can be distinguished from unbound

polynucleotide comprising the alternative recognition site. Binding reactions comprising either no glycoconjugate DNA ligand (control) or varying concentrations of the structural congener are incubated and electrophoresed. Glycoconjugate DNA ligands that selectively inhibit or displace binding of the predetermined DNA-binding protein to its recognition site will reduce the relative abundance of the bound complex comprising the predetermined DNA-binding protein and the polynucleotide comprising the predetermined recognition site (a retarded mobility band), preferably in a concentration-dependent relationship. Preferably, the glycoconjugate DNA ligand substantially will not displace or inhibit binding of the alternative DNA-binding protein(s) to the polynucleotide(s) comprising the cognate alternative recognition site(s) at the lowest concentrations of glycoconjugate DNA ligand which produces detectable inhibition of binding of the predetermined DNA-binding protein to its recognition site.

Novel sequence-selective glycoconjugate DNA ligands that bind preferentially to a specific DNA sequence can be identified by screening a library of glycogonjugate DNA ligands with a predetermined DNA sequence. For example, such a method generally comprises the steps of (1) generating a library of glycoconjugate DNA ligands of varying structural complexity by mixed chemical synthesis techniques known in the art, (2) attaching the library members to or synthesizing the library members directly on a solid phase support (e.g., beads comprising polystyrene crosslinked with 1 percent divinylbenzene), such that each unique ligand (library member) is present on discrete solid support or is otherwise spatially defined, (3) contacting the glycoconjugate DNA ligand library with a labeled (e.g., with a radiolabel or fluorescent label) polynucleotide comprising a predetermined sequence (e.g., a recognition sequence to which a predetermined sequence-selective DNA-binding protein binds), (4) identifying library members that selectively bind to (e.g., preferentially bind as compared to competitor random-sequence DNA, if included) the labeled polynucleotide, and (5) optionally determining the structure of the library member glycoconjugate

DNA ligand bound to the labeled polynucleotide. Solid phase oligosaccharide synthesis has been described (Danishefsky et al. Science 260: 1307, incorporated herein by reference) and may be used to generate combinatorial glycoconjugate libraries on solid substrates.

DNA sequences that bind to de novo synthesized glycoconjugate compounds that lack a predetermined DNA recognition site can be identified by screening a library of DNA sequences with a unique glycoconjugate ligand. Such methods generally comprise the steps of: (1) synthesizing a glycoconjugate ligand (e.g., according to a structural formulae provided herein), (2) attaching the glycoconjugate ligand to or synthesizing directly on a solid phase support, (3) contacting the glycoconjugate ligand with a library of polynucleotides comprising a variable region sequence of random, pseudorandom, or defined set sequences flanked by predetermined constant region of specific sequence(s) suitable for hybridizing to a PCR primer(s), such contacting is performed under suitable binding conditions (e.g., physiological conditions), (4) selecting the polynucleotides bound to the immobilized glycoconjugate ligand, (5) amplifying the selected polynucleotide sequences (e.g., by PCR), and (6) identifying the selected variable sequences by sequencing the PCR product(s).

Transcription Assays

Glycoconjugate DNA ligands can inhibit (or otherwise modify or modulate) the function of DNA-binding proteins that are transcription factors (e.g., NFAT, HNF-1, OCT-1/OAP) by interfering with their binding to their recognition DNA sequences, preferably without substantial interference with the function of other transcription factors or normal cell physiology. If the specificity of transcriptional inhibition is sufficiently high, such that transcription of genes dependent on one particular transcription factor (e.g., NFAT) are differentially inhibited to a sufficiently greater extent than genes having transcription dependent on essential or "housekeeping" transcription factors (e.g., Sp1, AP1), then the

differential inhibition of transcription may be exploited for therapeutic gene modulation, such as for medical treatment or veterinary use (e.g., for inducing or repressing transcription of selected transgenes under the transcriptional control of a transcription factor recognition sequence in transgenic animals).

Thus, the potency, sequence-selectivity, and efficacy of a glycoconjugate DNA ligand to differentially inhibit transcription driven by one particular transcription factor (or subgroup of transcription factors) can be readily evaluated using transcription assays.

Structural congeners which exhibit sequence-selective binding to a sequence present in a predetermined transcription factor recognition sequence can be further evaluated for their abilities to specifically modulate (e.g., inhibit) transcription of a polynucleotide sequence operably linked to a transcriptional regulatory sequence under the transcriptional influence of the predetermined transcription factor recognition sequence. In vitro transcription reactions may be used, if available, to ascertain the ability of a structural congener to modulate (e.g., inhibit) the transcription of a template under the transcriptional influence of a cis-linked predetermined transcription factor recognition sequence in the presence of the predetermined transcription factor (and other requisite transcription reaction components), wherein the modulation (inhibition) is determined relative to a control transcription reaction lacking the structural congener. Usually, one or more alternative transcription factors and polynucleotide sequences under the transcriptional influence of their cognate alternative recognition sequence(s) are assayed in parallel transcription reactions or in a single transcription reaction containing both the predetermined and the alternative DNA-binding proteins and recognition sequences.

Alternatively, or in combination with in vitro transcription assays, in vivo transcription assays may be performed. Eukaryotic or prokaryotic cells, preferably mammalian, insect, bacterial (e.g., *E. coli*), yeast, or plant cells, are employed as reporter host cells. Reporter host cells

comprise a reporter polynucleotide sequence (e.g., luciferase, β-galactosidase, IL-2, or other detectable sequence) under the transcriptional influence of an operably linked cis-acting transcription factor recognition sequence (e.g., an NFAT recognition site). Generally, the reporter host cells are stably or transiently transfected with one or more species of reporter polynucleotide according to any of the various methods and vectors known in the art. Preferably, the reporter host cells comprise a reporter polynucleotide under the transcriptional control of a predetermined transcription factor recognition sequence (e.g., NFAT site) and one or more alternative reporter polynucleotide(s) under the transcriptional control of alternative transcription factor recognition site(s) (e.g., Sp1, AP1 sites), wherein the alternative reporter polynucleotide(s) produce a transcribed signal (e.g., mRNA sequence, enzyme, etc.) that is distinguishable from the transcribed signal of the predetermined reporter polynucleotide.

Reporter host cells are cultured under suitable culture conditions for driving transcription of the linked reporter sequences under the influence of the linked transcription factor recognition sequence(s) (e.g., NFAT-dependent transcription is assayed in stimulated T lymphocyte lines). A glycoconjugate DNA ligand is added to the cell cultures in varying concentrations or dosages (i.e., for establishing a dose-response relationship) and the ability of the glycoconjugate DNA ligand to preferentially or specifically modulate (e.g., inhibit) the transcriptional activity of the predetermined reporter polynucleotide as compared to the alternative reporter polynucleotide(s) is determined. Structural congeners which preferentially or selectively inhibit transcription of the predetermined reporter polynucleotide as compared to the alternative reporter polynucleotide(s) are thereby identified as selective transcriptional antagonists. Preferably, such selective transcriptional antagonists do not produce substantial adverse effects (e.g., decreased cell viability, neoplastic transformation, expression of undesirable phenotypes) on the cultured cells at the lowest dose levels required for detectable

selective inhibition of transcription of the predetermined reporter polynucleotide. Generally, selective transcriptional antagonists are candidate pharmaceuticals for therapeutic transcriptional modulation of endogenous genes under the transcriptional influence of a cis-acting predetermined transcription factor recognition sequence (e.g., NFAT-dependent genes).

Selective transcriptional antagonists are administered to nonhuman animals (e.g., mice, rats, rabbits, hamsters, nonhuman primates) to determine safety, efficacy, and selectivity of transcriptional modulation of the desired endogenous genes in the desired cell type and organs. The therapeutic effect of treating disease conditions by such gene modulation is evaluated and maximum tolerated dose, LD_{50}, and a dose-response curve is generated to calibrate efficacious dose levels and routes of administration. In one variation, calicheamicin MG is employed to selectively inhibit transcription of NFAT-dependent genes and thereby inhibit T cell activation and effect immunosuppression of cell-mediated immune response (e.g., allograft rejection) in vivo.

Commercial Research Reagent

In addition to their utility as therapeutic agents for treating human diseases, glycoconjugate DNA ligands, such as the use of calicheamicin MG and related structural congeners for selective antagonism of NFAT-dependent gene transcription and immunosuppression, the glycoconjugate DNA ligands will find use as commercial reagents.

For example and not limitation, NFAT can be purified in preparative quantities by methods wherein a purification step involves affinity chromatography of a sample containing NFAT (e.g., an extract from activated T lymphocytes) on a matrix containing DNA sequences comprising NFAT binding sites to which calicheamicin MG selectively binds. The sample containing NFAT is contacted with the affinity matrix under binding conditions permitting the binding of NF-At to its DNA recognition site(s) on the matrix DNA. Unbound sample is removed, and the matrix is

optionally washed with a suitable wash buffer. As an enrichment step, a solution comprising an effective concentration (or concentration gradient) of calicheamicin MG is applied to the bound matrix to selectively displace, and thus elute, NFAT from its recognition site(s) on the matrix DNA without eluting all DNA-binding proteins bound to the DNA matrix.. The eluate is enriched for NFAT.

Alternatively, calicheamicin MG may be used for titrating NFAT binding in gel shift assays, DNA footprinting, nuclease-protection assays, and the like. Many such assays may have diagnostic as well as research applications.

It is envisioned that a significant commercial market will develop for calicheamicin MG as a research reagent (much as restriction endonucleases are widely sold for research and diagnostic uses), particularly in view of the desirability of using purified NFAT for transcription assays, binding assays, and the like for large-scale drug screening programs for novel immunosuppressants, and the like.

Pharmaceutical Compositions

Glycoconjugate DNA ligands, such as the non-cytotoxic calicheamicin-related compounds of Structures I, II, and III, and calicheamicin MG, can be used as pharmaceuticals for effecting therapeutic and/or prophylactic transcriptional modulation of predetermined genes in vivo. For example, calicheamicin MG can act as an immunosuppressant by selectively inhibiting NFAT-dependent transcription of T cell activation genes (e.g., IL-2). Calicheamicin MG and other structural congeners of calicheamicin and esperimicin may exhibit selective agonism or antagonism of the expression of other genes, for example proto-oncogenes, oncogenes, tumor suppressor genes, genetic disease alleles, pathogen-related genes, and transgenes

The preferred pharmaceutical compositions of the present invention comprise a therapeutically or prophylactically effective dose of at least one glycoconjugate DNA ligand, such as calicheamicin MG.

Pharmaceutical compositions comprising a glycoconjugate

DNA ligand of the present invention are useful for topical and parenteral administration, i.e., subcutaneously, intramuscularly or intravenously. The finding that glycoconjugate DNA ligands possess sequence-specific transcriptional inhibition in vitro as well as in vivo without substantial cytotoxicity indicates that calicheamicin-related glycoconjugate DNA ligands are suitable for pharmaceutical use. The glycoconjugate DNA ligands of the invention are suitable for administration to mammals, including human patients and veterinary patients.

The compositions for parenteral administration will commonly comprise a solution of a glycoconjugate DNA ligand or a cocktail thereof dissolved in an acceptable carrier, preferably an aqueous carrier. It is often preferable to include in the carrier a hydrophobic base (e.g., polyethylene glycol, Tween 20). A variety of aqueous carriers can be used, e.g., water, buffered water, 0.4% saline, 0.3% glycine and the like. These solutions are sterile and generally free of particulate matter. These compositions may be sterilized by conventional, well known sterilization techniques. The compositions may contain pharmaceutically acceptable auxiliary substances as required to approximate physiological conditions such as pH adjusting and buffering agents, toxicity adjusting agents and the like, for example sodium acetate, sodium chloride, potassium chloride, calcium chloride, sodium lactate, etc. The concentration of the glycoconjugate DNA ligand(s) in these formulations can vary widely, i.e., from less than about 1 nM, usually at least about 0.1mM to as much as 100 mM and will be selected primarily based on fluid volumes, viscosities, etc., in accordance with the particular mode of administration selected. Most usually, the glycoconjugate DNA ligand is present at a concentration of 0.1 μM to 10 mM. For example, a typical formulation for intravenous injection comprises a sterile solution of a non-cytotoxic glycoconjugate DNA ligand at a concentration of 5 mM in Ringer's solution. A hydrophobic vehicle may be used, or an aqueous vehicle comprising a detergent or other lipophilic agent (e.g., Tween, NP-40, PEG); alternatively, the glycoconjugate DNA ligand(s) may be administered as a suspension in an aqueous

carrier, or as an emulsion.

Thus, a typical pharmaceutical composition for intramuscular injection could be made up to contain 1 ml sterile buffered water, and about 1-100 mg of glycoconjugate DNA ligand. A typical composition for intravenous infusion can be made up to contain 250 ml of sterile Ringer's solution, and about 100-1000 mg of glycoconjugate DNA ligand(s). Actual methods for preparing parenterally administrable compositions will be known or apparent to those skilled in the art and are described in more detail in, for example, Remington's Pharmaceutical Science, 15th Ed., Mack Publishing Company, Easton, Pennsylvania (1980), which is incorporated herein by reference. A typical pharmaceutical composition for topical application can be made with suitable dermal ointments, creams, lotions, ophthalmic ointments and solutions, respiratory aerosols, and other excipients. Excipients should be chemically compatible with the glycoconjugate DNA ligands that are the active ingredient(s) of the preparation, and generally should not increase decomposition, denaturation, or aggregation of active ingredient(s). Frequently, excipients will have lipophilic components such as oils and lipid emulsions.

The transcriptional antagonist glycoconjugate DNA ligands of this invention can be lyophilized for storage and reconstituted in a suitable carrier prior to use. It will be appreciated by those skilled in the art that lyophilization and reconstitution can lead to varying degrees of activity loss, and that use levels may have to be adjusted to compensate.

The compositions containing the present glycoconjugate DNA ligands (e.g., selective transcription antagonists) or cocktails thereof can be administered for prophylactic and/or therapeutic treatments. In therapeutic application, compositions are administered to a patient already affected by the particular disease, in an amount sufficient to cure or at least partially arrest the condition and its complications my modulating expression of one or more predetermined genes. An amount adequate to accomplish this is defined as a "therapeutically effective dose" or "efficacious dose." Amounts effective for

this use will depend upon the severity of the condition, the general state of the patient, and the route of administration, but generally range from about 1 mg to about 10g of glycoconjugate DNA ligand per dose, with dosages of from 10 mg to 2000 mg per patient being more commonly used. For example, for treating acute tissue graft rejection reactions, about 10 to 1000 mg of a calicheamicin MG or a congener may be administered systemically by intravenous infusion.

In prophylactic applications, compositions containing the glycoconjugate DNA ligands or cocktails thereof are administered to a patient not already in a disease state to enhance the patient's resistance or to retard the progression of disease. Such an amount is defined to be a "prophylactically effective dose." In this use, the precise amounts again depend upon the patient's state of health and general level of immunity, but generally range from 1 mg to 10 g per dose, especially 10 to 1000 mg per patient. A typical formulation of a glycoconjugate DNA ligand used as a selective transcriptional antagonist, such as calicheamicin MG, will contain between about 25 and 250 mg of the glycoconjugate in a unit dosage form.

Single or multiple administrations of the compositions can be carried out with dose levels and dosing pattern being selected by the treating physician. In any event, the pharmaceutical formulations should provide a quantity of the glycoconjugate DNA ligand(s) of this invention sufficient to effectively treat the patient. Typically, at least one species of glycoconjugate DNA ligand is administered as the sole active ingredient, or in combination with one or more other active ingredients. In general for treatment of gene-expression related disease (e.g, autoimmunity, neoplasia, senescence, congenital genetic disease), a suitable effective dose of the glycoconjugate transcriptional modulator will be in the range of 0.01 to 1000 milligram (mg) per kilogram (kg) of body weight of recipient per day, preferably in the range of 1 to 100 mg per kg of body weight per day. The desired dosage is preferably presented in one, two, three, four or more subdoses administered at appropriate intervals throughout the day. These subdoses can be administered

as unit dosage forms, for example, containing 5 to 10,000 mg, preferably 10 to 1000 mg of active ingredient per unit dosage form.

Once detectable improvement of the patient's conditions has occurred, a maintenance dose is administered if necessary. Subsequently, the dosage or the frequency of administration, or both, can be reduced, as a function of the symptoms, to a level at which the improved condition is retained. When the symptoms have been alleviated to the desired level, treatment can cease. Patients can, however, require intermittent treatment on a long-term basis upon any recurrence of the disease symptoms or as a prophylactic measure to prevent disease symptom recurrence. The composition used in these therapies can be in a variety of forms. These include, for example, solid, semi-solid and liquid dosage forms, such as tablets, pills, powders, liquid solutions or suspensions, liposome preparations, injectable and infusible solutions. The preferred form depends on the intended mode of administration and therapeutic application. Typically, a sterile solution of a glycoconjugate DNA ligand in an aqueous solvent (e.g., saline) will be administered intravenously. The compositions also preferably include conventional pharmaceutically acceptable carriers and adjuvants which are known to those of skill in the art. See, e.g., Remington's Pharmaceutical Sciences, Mack Publishing Co.: Easton, PA, 17th Ed. (1985). Generally, administration will be by oral or parenteral (including subcutaneous, intramuscular, intravenous, and intradermal) routes, or by topical application or infusion into a body cavity, or as a bathing solution for tissues during surgery.

For solid compositions, conventional nontoxic solid carriers can be used which include, for example, pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, sodium saccharin, talcum, cellulose, glucose, sucrose, magnesium carbonate, and the like. For oral administration, a pharmaceutically acceptable nontoxic composition is formed by incorporating any of the normally employed excipients, such as those carriers previously listed, and generally 0.001-95% of

active ingredient, preferably about 20%. It should, of course, be understood that the methods of this invention can be used in combination with other agents that have gene expression modulatory activity and/or other immunosuppressants (e.g., cyclosporin A, FK506).

While it is possible to administer the active ingredient of this invention alone, it is believed preferable to present it as part of a pharmaceutical formulation. The formulations of the present invention comprise at least one compound of this invention in a therapeutically or pharmaceutically effective dose together with one or more pharmaceutically or therapeutically acceptable carriers and optionally other therapeutic ingredients. Various considerations are described, e.g., in Gilman et al. (eds) (1990) Goodman and Gilman's: The Pharmacological Bases of Therapeutics, 8th Ed., Pergamon Press; and Remington's supra, each of which is hereby incorporated herein by reference. Methods for administration are discussed therein, e.g., for oral, intravenous, intraperitoneal, or intramuscular administration, and others. Pharmaceutically acceptable carriers will include water, saline, buffers, and other compounds described, e.g., in the Merck Index, Merck & Co., Rahway, NJ, incorporated herein by reference.

A more recently devised approach for parenteral administration employs the implantation of a slow-release or sustained-release system, such that a constant level of dosage is maintained. See, e.g., U.S. Patent No. 3,710,795, which is incorporated herein by reference. Glycoconjugate DNA ligands may be administered by transdermal patch (e.g., iontophoretic transfer) for local or systemic application.

Kits can also be supplied for use with the subject glycoconjugates for use in the protection against or therapy for a disease. Thus, the subject composition of the present invention may be provided, usually in a lyophilized form or aqueous solution in a container, either alone or in conjunction with additional non-cytotoxic glycoconjugate DNA ligands of the desired type. The non-cytotoxic glycoconjugate DNA ligands are included in the kits with buffers, such as Tris, phosphate,

carbonate, etc., stabilizers, biocides, inert proteins, e.g., serum albumin, or the like, and a set of instructions for use. Generally, these materials will be present in less than about 5% wt. based on the amount of glycoconjugate, and usually present in total amount of at least about 0.001% based again on the concentration. Frequently, it will be desirable to include an inert extender or excipient to dilute the active ingredients, where the excipient may be present in from about 1 to 99.999% wt. of the total composition.

In Vitro and Research Administration

In another aspect of the invention, calicheamicin-related glycoconjugate DNA ligand(s) of the invention are employed to modulate the expression of naturally-occurring genes or other polynucleotide sequences under the transcriptional control of a predetermined transcription regulatory element comprising a transcription factor binding site (e.g., an NFAT site). Transgenes, homologous recombination constructs, and episomal expression systems (e.g., viral-based expression vectors) comprising a polynucleotide sequence under the transcriptional control of one or more transcription factor binding site linked to a promoter will be made by those of skill in the art according to methods and guidance available in the art, as will transformed cells and transgenic nonhuman animals harboring such polynucleotide constructs. The transcriptional antagonist (or agonist) glycoconjugate DNA ligands may be used to modulate the transcription of predetermined transcription factor-regulated polynucleotide sequences in cell cultures (e.g., ES cells) and in intact animals, particularly in transgenic animals wherein a transgene comprises one or more predetermined transcription factor recognition sites as transcriptional regulatory sequences. For transformed or transgenic cell cultures, a dose-response curve is generated by titrating transcription rate of the recognition site-controlled polynucleotide sequence against increasing concentrations of selectively binding glycoconjugate DNA ligand, which will reduce the transcription rate reliant upon binding of the predetermined

transcription factor to its recognition site. Similar dose-response titration can be performed in transgenic animals, such as transgenic mice, harboring a predetermined transcription factor recognition site-controlled transgene sequence.

The broad scope of this invention is best understood with reference to the following examples, which are not intended to limit the invention in any manner. The following examples are offered by way of illustration, not by way of limitation.

EXPERIMENTAL EXAMPLES

Overview

The ordered regulation of gene expression that determines specific developmental programs or cellular responses to physiologic stimuli is dependent on sequence-specific DNA-protein interactions. This requisite specificity suggests a means to control phenotypic responses through the use of ligands that (i) bind DNA with a high degree of sequence specificity, (ii) antagonize DNA-protein interactions either by inhibition of DNA-protein complex formation or by displacement of preformed complexes, and (iii) function *in vivo*. Several types of compounds have been shown to satisfy some of these criteria, including oligodeoxynucleotides involved in triple helix formation, peptide nucleic acids, and other nonintercalating DNA ligands such as netropsin or distamycin A. However, none of these adequately satisfy all the requirements for specific and biologically useful transcriptional antagonists.

The enediyne compound calicheamicin $\gamma 1^I$ (CLM, Fig. 1), representative of a novel class of heterocyclic glycoconjugate DNA ligands, binds to and cleaves DNA at specific four base pair sequence. Affinity cleavage experiments performed by Zein et al. demonstrate that CLM preferentially cleaves DNA at cytosine-containing homopyrimidine tracts, preferentially TCCT. Additional studies, however, show that in the absence of a TCCT sequence the preferential site of cleavage is TTTT despite the presence of a cytosine-containing sequence TCCC. Thus, although CLM does exhibit preferential binding at the four base pair level

suggestive of base specific interactions (i.e. direct readout), sequence-dependent DNA structural polymorphism and conformational flexibility are also likely to significantly influence specificity (i.e. indirect readout). Comparisons of the DNA cleavage specificity of CLM derivatives indicate that the polysaccharide/aromatic component is largely responsible for sequence-specific DNA binding.

To demonstrate sequence-specific DNA-ligand interactions, the methyl glycoside derivative of the aryl tetrasaccharide of calicheamicin CLM (hereafter referred to as CLM-MG or calicheamicin MG) which lacks the enediyne component and therefore does not cleave DNA was synthesized.

Synthesis of Calicheamicin MG

Calicheamicin MG was synthesized according to the general method of Halcomb et al. (1992) Ang. Chem., Int. Ed. Engl. 31: 338 and Halcomb et al. (1992) Ang. Chem. 104: 314.

Analysis of DNA-Protein Interactions

DNA footprinting analysis indicated that CLM-MG exhibits DNA sequence specific binding that is similar to the parent compound, CLM. To determine if CLM-MG binding was capable of interfering with DNA-protein interactions, the affect of CLM-MG on the binding of several distinct DNA binding proteins *in vitro* was measured using a electrophoretic mobility shift assay. The transcription factors NFAT, AP1, Oct-1 and Oct-1 associated protein (OAP), NFkB, HNF-1, and Sp1 represent multimeric protein complexes that bind to specific DNA sequences found in the 5' regulatory regions of a variety of inducible, constitutive, and tissue-specific genes. These protein complexes are representative members of several distinct classes of transcription factors, including the basic-leucine zipper or bZIP class (AP1), the rel/dorsal class (NF-kB), the homeodomain class which contains a helix-turn-helix motif (NHF1, Oct-1), and the zinc finger class (Sp1). The constituent proteins of NFAT have not been completely characterized, but appear to consist of a protein sharing some characteristics with AP1 and an unidentified DNA-binding protein that undergoes cytoplasmic to nuclear

translocation upon T cell antigen receptor stimulation. Predicted CLM binding sites are present within four of the six DNA sequences to which these transcription factors bind (Fig. 2). Furthermore, only the NFAT recognition sequence contains the presumed preferential CLM recognition site TCCT.

Nuclear extracts containing the various DNA binding activities were prepared as previously described (Fiering et al. (1990) Genes Devel. 4: 1823, incorporated herein by reference) from TAg Jurkat cells (a derivative of the human T cell leukemia cell line Jurkat transfected with the SV40 large T antigen (Northrop et al. (1993) J. Biol. Chem. 268: 2917) stimulated with 2 μM ionomycin plus 20 ng/ml 12-O-tetradecanoyl-phorbol-13-acetate (TPA) for two hours (NFAT, AP1, Oct1, NF-kB), HeLa cells stimulated with 20 ng/ml TPA for 2 hours (AP1, Oct1, NF-kB, Sp1), or MDCK cells (HNF-1). Cells were maintained in RPMI-1640 supplemented with 10% (vol/vol) heat-inactivated fetal calf serum, 100 U/ml penicillin, 100 μg/ml streptomycin, 2 mM L-glutamine and 50 μM β-mercaptoethanol (complete medium) in a 5% CO_2/95% air humidified atmosphere. CLM-MG was solubilized in 10% EtOH or DMSO. Electrophoretic mobility shift assays were performed by incubating varying concentrations of CLM-MG or the appropriate buffer control with 0.1 ng DNA probe (approx. 20,000 cpm, see below) for 15 minutes at room temperature, followed by the addition of 2 μL (5-10 μg) nuclear extract, in a final volume of 15 μL. The incubation was continued for 45 minutes at room temperature. The binding buffer consisted of 10 mM Tris-Cl (pH 7.6), 80 mM NaCl, 1 mM EDTA, 0.5 mM dithiothreitol, 5% glycerol, and 2.5 μg poly d(I-C). Samples were subsequently loaded onto a 4% non-denaturing polyacrylamide gel in 0.5X TBE and electrophoresed at 180 volts for approximately 1.5-2 hours at room temperature. DNA probes (listed below) were labelled with ^{32}P-γ-ATP and polynucleotide kinase, followed by filling-in 5' overhang with dNTPs and the Klenow fragment of DNA polymerase (as necessary). Alternatively, labelling was performed by filling-in 5' overhangs with ^{32}P-α-dCTP, dATP, dTTP, dGTP and the Klenow fragment of DNA polymerase. Protein-DNA complex formation was measured directly using a radioanalytic imaging system (Ambis

Imaging, Inc.). The following oligonucleotides were utilized as probes for the gel shift assays:

NFAT	5'	GATCTAAGGAGGAAAAACTGTTTCATG
	3'	ATTCCTCCTTTTTGACAAAGTACCTAG
AP1	5'	TCGAGTGACTCAGCGCG
	3'	CACTGAGTCGCGCAGCT
NFkB	5'	AGGGATTTCAC
	3'	TCCCTAAAGTG
Oct1	5'	TGTAATATGTAAAACATTT
	3'	ATTATACATTTTGTAAAAC
Oct1/OAP	5'	TTTGAAAATAGTGTAATATGTAAAACAT
	3'	CTTTTATACACATTATACATTTTGTAAAA
HNF-1	5'	CAAACTGTCAAATATTAACTAAAGGGAG
	3'	GTTTGACAGTTTATAATTGATTTCCCTC
Sp1	5'	TCGAGGGGCGGGGC
	3'	CCCCGCCCCGAGCT

The presence of CLM-MG in DNA-protein binding reactions resulted in the complete inhibition of NFAT complex formation at 100 μM CLM-MG. In contrast, there was no significant inhibition of AP1 or Sp1 complex formation at concentrations of CLM-MG which significantly inhibited NFAT-DNA complex formation and minimal inhibition at concentrations as high as 100 μM. NF-kB, Oct-1, and HNF-1 showed intermediate levels of inhibition of DNA-protein complex formation by CLM-MG (Fig. 3). The greater sensitivity of OAP to inhibition by CLM-MG as compared to Oct-1 is likely a reflection of the dependence of OAP binding on Oct-1 binding. These results indicate that CLM-MG is capable of inhibiting DNA-protein complex formation and are consistent with the result predicted based on the presence or absence of CLM binding sites within the transcription factor recognition sequences.

Studies of the kinetics of CLM-MG inhibition of DNA-protein interaction showed that preincubation of CLM-MG with DNA was not necessary for complete inhibition of NFAT-DNA complex formation. Furthermore, CLM-MG rapidly displaced preformed NFAT-DNA complexes, with 80% inhibition of NFAT-DNA complex formation observed within 15 seconds of the addition of CLM-MG followed by

separation of bound and unbound DNA by non-denaturing acrylamide gel electrophoresis. In contrast, addition of excess unlabeled DNA probe resulted in far less dissociation of the DNA-protein complex. The rapid dissociation of preformed DNA-protein complexes by CLM-MG indicates that sensitivity to inhibition by CLM-MG is not dependent on differences in DNA-protein complex dissociation rates. In addition, the similarity between the rate of dissociation of preformed NFAT-DNA induced by CLM-MG at room temperature and the apparent rate at 4°C is consistent with the rate constant for the association of CLM-MG and DNA being diffusion-collision limited.

The differential sensitivity of the NFAT-DNA, HNF1-DNA, and AP1-DNA complexes to disruption by CLM-MG is shown in Fig. 4, demonstrating that NFAT-DNA complexes are most sensitive to disruption by CLM-MG.

To provide further evidence that CLM-MG inhibits DNA-protein complex formation by binding to DNA in a sequence specific manner, alterations were made in the putative CLM-MG binding sites within the NFAT recognition sequence (Fig. 5). Substitution of thymidine for guanine in the first guanine dinucleotide resulted in decreased sensitivity of the NFAT complex to inhibition by CLM-MG (Figs. 5 and 6). Alterations in the second guanine dinucleotide eliminated NFAT binding, consistent with the identification of these nucleotides as contact sites by methylation interference studies. The presence of overlapping CLM-MG binding sites within the NFAT recognition sequence gives rise to the prediction that increasing the distance between these sites may increase sensitivity to CLM-MG, since overlapping sites may result in decreased occupation of the site more directly involved in protein contacts. Comparison between the wild-type NFAT recognition sequence and an altered sequence containing an inserted adenine nucleotide shows that binding of NFAT to the later sequence is more sensitive to inhibition by CLM-MG (Fig. 6). Cross-competition studies between the wild-type and altered DNA sequences demonstrate that NFAT binds to the altered DNA sequences with approximately equal affinity as the wild-type sequence. Thus, the extent to which

NFAT-DNA complex formation is inhibited by CLM-MG is dependent on the presence of CLM-MG binding sites, providing further evidence demonstrating that CLM-MG functions to inhibit DNA-protein interaction by binding to DNA in a sequence specific manner.

The converse experiment of generating CLM-MG sensitivity by introduction of a CLM-MG binding site into a DNA protein recognition sequence that is not susceptible to inhibition to CLM-MG was performed by altering the AP1 recognition sequence. A CLM-MG binding site was introduced immediately adjacent (3') to the AP1 recognition sequence ACTGTCA. Comparison between the wild type AP1 sequence and the altered sequence shows that introduction of a CLM-MG binding site increases the sensitivity of AP1-DNA complexes to CLM-MG. However, these changes also resulted in less binding by AP1 and therefore the possibility that the increased sensitivity to CLM-MG is a result of a decreased protein-DNA affinity cannot be excluded.

Transcriptional Assays

Finally, to determine whether the sequence-specific inhibition of DNA-protein interactions by CLM-MG observed <u>in vitro</u> could be manifest <u>in vivo</u>, the affect of CLM-MG on the expression of the secreted alkaline phosphatase reporter gene directed by multimerized NFAT, AP1, NFkB or Oct1/OAP recognition sequences located upstream of the IL-2 minimal promoter was measured in transient transfection assays.

Reporter constructs driving expression of an alkaline phosphatase gene under transcriptional control of cis-linked sequences were made for transient expression assays. NFAT-SX contains three copies of the NFAT recognition sequence (-286 to -257 of the human IL-2 enhancer). AP1-SX contains five copies of the human metallothionein promoter/enhancer AP-1 binding site (5'-TGACTCAGCGC-3'). NFkB-SX contains three copies of the of the murine κ light chain NFkB binding site. Oct1/OAP-SX contains four copies of the Oct1/OAP site (ARRE1; -93 to -63 of the human IL-2 enhancer).

TAg Jurkat cells were transiently transfected with 4

μg of the indicated reporter plasmid by electroporation in complete media (Bio-Rad GenePulser; 960 μF, 250 V, 0.4 cm cuvette width). Cells were harvested after approximately 24 hours and aliquoted into 96-well flat bottom microtitre plates (2×10^5 cells per well in 100 μl of complete medium). Varying concentrations of CLM-MG or the control buffer were added to duplicate wells in a 5 or 10 μl volume. After 60 minutes at 37°C, ionomycin and TPA were added in 100 μl complete medium to final concentrations of 1 μM and 20 ng/ml, respectively. Secreted alkaline phosphatase activity was measured after 12-16 hours as described in Clipstone N and Crabtree GR (1992) Nature 357: 695, incorporated herein by reference.

Expression of each of these reporter constructs requires stimulation of the transfected cells by cross-linking the T cell antigen receptor or by treatment with ionomycin and/or TPA. CLM-MG markedly inhibited the induced expression of the reporter gene driven by the NFAT-IL2 promoter with a half-maximal inhibitory concentration of 50-75 μM. In contrast, expression of the reporter gene driven by the AP1-IL2 promoter showed significantly less inhibition by CLM-MG, with a half-maximal inhibitory concentration of 225 μM. The NFkB-IL2 and Oct1/OAP-IL2 promoters were inhibited to an intermediate level (Fig. 7). These results reflect in vivo the same specificity of inhibition of DNA-protein complex formation by CLM-MG that was observed in vitro and are consistent with CLM-MG inhibiting gene expression in vivo by inhibiting protein-DNA complex formation in a sequence specific manner. Although in these experiments CLM-MG was added to cells 60 minutes prior to stimulation, in other experiments significant inhibition was seen if CLM-MG was added only 5 minutes prior to stimulation. This is consistent with the rapid rate in which CLM-MG binds DNA in vitro. Interestingly, addition of CLM-MG 15 minutes after stimulation resulted in minimal inhibition of reporter gene expression, indicating that once transcription is initiated either (i) enhancer binding proteins are no longer required for transcription, (ii) enhancer binding proteins are no longer susceptible to displacement by CLM-MG, (iii) CLM-MG acts by inhibiting the formation of DNA-protein

complexes but cannot displace preformed complexes, or (iv) alternate mechanisms are involved.

Transcription factors typically form interactions with DNA along the major groove. CLM, in contrast, appears to occupy the minor groove. Assuming the DNA-protein interactions that are sensitive to inhibition by CLM-MG depend primarily on major groove contacts, the rapid displacement of preformed DNA-protein complexes by CLM-MG is unlikely to result from direct competition for a shared binding site. Rather, CLM-MG binding in the minor groove may either induce a DNA conformation that is incompatible with protein binding along the major groove or render the specific region of DNA more rigid, thereby prohibiting DNA conformational changes that may be required for effective binding protein. This hypothetical binding interaction contrasts with that of the minor groove ligand netropsin, which binds triple helix DNA without displacing the major groove-bound third strand, or 1-methylimidazole-2-carboxamide netropsin, which can bind simultaneously with the major groove-binding protein GCN4(226-281) at a common binding site. Thus, targeting the minor groove, which appears to be accessible to small ligands in the presence of protein interacting along the major groove, may be a useful strategy in the design of DNA ligands capable of conferring sufficient conformational constraints upon DNA to antagonize DNA-protein interactions and thereby inhibit gene expression. As with most interactions between complex molecules, the specificity of the interaction between CLM-MG and DNA most likely results from contributions of specific hydrogen bond configurations substituting for the spine of hydration present in the minor groove of DNA and close van der Waals contacts between CLM-MG and appropriate constituents of DNA present along the floor and walls of the minor groove. Electrostatic interactions between CLM-MG and DNA likely contribute less to the specificity of the interaction due to the lack of charged groups in CLM-MG. Further understanding of the precise determinants of sequence-specific binding by CLM-MG and of the exact mechanism by which CLM-MG inhibits DNA-protein interactions and displaces preformed DNA-protein complexes can be determined by direct structural analyses

of CLM binding sequences and CLM-DNA complexes. The results presented here demonstrate that CLM-MG exhibits many of the properties required for biologically useful, sequence-specific transcriptional antagonists and suggest that CLM-MG and other glycoconjugate DNA ligands may serve as lead compounds in the development of novel DNA-targeted therapeutic agents and biologic probes, among many other uses.

Although the present invention has been described in some detail by way of illustration for purposes of clarity of understanding, it will be apparent that certain changes and modifications may be practiced within the scope of the claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Ho, Steffan N.
Schreiber, Stuart L.
Danishefsky, Samuel
Crabtree, Gerald R.

(ii) TITLE OF INVENTION: SEQUENCE-SPECIFIC GLYCOCONJUGATE TRANSCRIPTIONAL ANTAGONISTS

(iii) NUMBER OF SEQUENCES: 31

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Townsend and Townsend Khourie and Crew
(B) STREET: 379 Lytton Avenue
(C) CITY: Palo Alto
(D) STATE: California
(E) COUNTRY: US
(F) ZIP: 94301

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/109,271
(B) FILING DATE: 18-AUG-1993
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Smith, Willaim M
(B) REGISTRATION NUMBER: 30,223
(C) REFERENCE/DOCKET NUMBER: 5490A-210

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (415) 326-2400
(B) TELEFAX: (415) 326-2422

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

AAGGAGGAAA AACTGTTTCA T 21

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GGGGCGGGGC 10

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GTGACTCAGC GCG 13

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

AAGAGGAAAA A 11

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GAAAGGAGGA AAAACTGTTT 20

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CCAAAGAGGA AAATTTGTTT 20

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CAGAAGAGGA AAAATGAAGG 20

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

TCCAGGAGAA AAAATGCCTC 20

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

AAAACTTGTG AAAATACGTA 20

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TAAAGGAGAG AACACCAGCT 20

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GCAGGGTGGG AAAGGCCTTT 20

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GATCTAAGGA GGAAAAACTG TTTCATG 27

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GATCCATGAA ACAGTTTTTC CTCCTTA 27

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

TCGAGTGACT CAGCGCG 17

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

TCGACGCGCT GAGTCAC 17

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

AGGGATTTCA C 11

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GTGAAATCCC T 11

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

TGTAATATGT AAAACATTT 19

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

CAAAATGTTT TACATATTA 19

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

TTTGAAAATA GTGTAATATG TAAAACAT 28

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

AAAATGTTTT ACATATTACA CATATTTTC 29

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CAAACTGTCA AATATTAACT AAAGGGAG 28

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

CTCCCTTTAG TTAATATTTG ACAGTTTG 28

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

TCGAGGGGCG GGGC 14

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

TCGAGCCCCG CCCC 14

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

TGACTCAGCG C 11

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

TGTAATATGT AAAACATTTT G 21

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 37 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GATCTTTGAA AATATGTGTA ATATGTAAAA CATTTTG 37

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GAAGGAGGAA AAA 13

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 13 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GAAGTAGGAA AAA 13

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

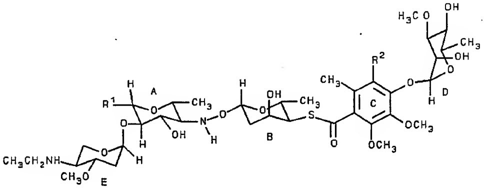
(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

GAAGGAAGGA AAAA 14

Claims

1. A composition of a glycoconjugate DNA ligand lacking DNA strand cleavage activity, comprising a compound according to the structure



wherein R_1 is selected from the group consisting of hydrogen, hydroxy, lower alkyl, lower alkoxy, aryloxy, and alkanol; and R_2 is selected from the group consisting of hydrogen, lower alkyl, lower alkoxy, iodo, bromo, fluoro, and chloro.

2. A composition comprising a glycoconjugate DNA ligand of claim 1, wherein R_1 is selected from the group consisting of methoxy, ethoxy, propyloxy, methyl, ethyl, and hydrogen.

3. A composition of claim 2, wherein R_2 is iodo, bromo, chloro, or fluoro.

4. A composition of claim 1, wherein the glycoconjugate DNA ligand binds to DNA in a sequence-selective manner with preferential binding to the tetranucleotide sequences TCCT and TCTC.

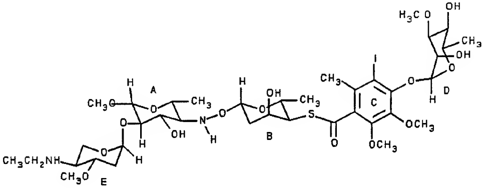
5. A composition of claim 4, wherein the glycoconjugate DNA ligand preferentially binds to an NFAT recognition sequence as compared to an AP1 or Sp1 recognition sequence.

6. A composition of claim 5, wherein the glycoconjugate DNA ligand inhibits the formation of an NFAT-DNA complex between NFAT and an NFAT recognition sequence or displaces NFAT from a

preformed NFAT-DNA complex.

7. A composition of claim 6, wherein the glycoconjugate DNA ligand is calicheamicin MG

8. A glycoconjugate DNA ligand having the structure:



9. A glycoconjugate DNA ligand of claim 8 which is labeled with a radioisotope or spin label.

10. A glycoconjugate DNA ligand of claim 8 noncovalently bound to a polynucleotide comprising a NFAT recognition site sequence forming a glycoconjugate-liganded NFAT recognition site.

11. A glycoconjugate-liganded NFAT recognition site comprising calicheamicin MG bound to a polynucleotide comprising a NFAT recognition site sequence.

12. A glycoconjugate-liganded NFAT recognition site of claim 11, wherein the polynucleotide is present in a mammalian cell.

13. A glycoconjugate-liganded NFAT recognition site of claim 12, wherein the polynucleotide present in said mammalian cell is replicable.

14. A method for modulating the transcriptional activity of polynucleotide sequences under the transcriptional influence

of an operably linked cis-acting sequence to which a DNA-binding protein binds, comprising the step of:

administering a sequence-selective glycoconjugate DNA ligand to a cell or to an in vitro transcription reaction wherein the glycoconjugate DNA ligand binds to said operably linked cis-acting sequence and alters binding of said DNA-binding protein to said operably linked cis-acting sequence, thereby modulating transcription of the polynucleotides sequence under the transcriptional influence of said operably linked cis-acting sequence.

15. A method of claim 14, wherein the operably linked cis-acting sequence comprises a transcription factor recognition site and the DNA-binding protein is a transcription factor.

16. A method of claim 15, wherein the transcription factor recognition site is an NF-AT recognition site and the DNA-binding protein comprises NF-AT.

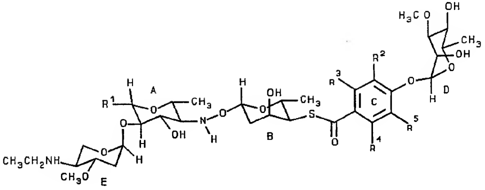
17. A method of claim 16, wherein the NF-AT recognition site comprises the polypeptide sequence spanning -286 to -257 upstream of the human IL-2 gene.

18. A method for producing immunosuppression in a human patient, comprising administering a prophylactically or therapeutically effective dose of a glycoconjugate DNA ligand that selectively binds to a NF-AT recognition site and inhibits NF-AT-dependent transcription, thereby modulating expression of a NF-AT-dependent gene in a T lymphocyte and inhibiting T lymphocyte activation.

19. A method for identifying a sequence-selective glycoconjugate DNA ligand, comprising the steps of:

administering to a binding reaction comprising a predetermined DNA-binding protein and a polynucleotide comprising a recognition site for said predetermined DNA-binding protein a glycoconjugate DNA ligand having a structure according to the

formula:



where R_1 is hydrogen, halogen, hydroxy, lower alkyl, lower alkoxy, or aryloxy; R_2 is a halogen, hydroxy, or lower alkyl; R_3 is lower alkyl, lower alkoxy, or hydrogen; and R_4 and R_5 are independently selected from lower alkyl, lower alkoxy, or hydrogen; and

detecting the ability of the glycoconjugate DNA ligand to inhibit binding of the predetermined DNA-binding protein to the recognition site, and;

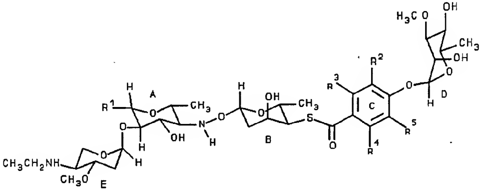
identifying a glycoconjugate DNA ligand which inhibits said binding of said predetermined DNA-binding protein to said recognition sequence as a sequence-selective glycoconjugate DNA ligand.

20. A method of claim 19, wherein at least one alternative DNA-binding protein and a polynucleotide comprising a recognition site to which the alternative DNA-binding protein binds are included in the binding reaction and wherein the step of detecting comprises determining the differential inhibition of binding of said predetermined DNA-binding protein to its recognition site as compared to an inhibition of binding of said alternative DNA-binding protein to its recognition site.

21. A method for identifying glycoconjugate DNA ligands which are differential transcriptional agonists or antagonists, comprising the steps of:

administering to a transcription assay comprising a reporter polynucleotide sequence under the transcriptional

influence of a operably linked cis-acting recognition site for a predetermined transcription factor a glycoconjugate DNA ligand having a structure according to the formula:



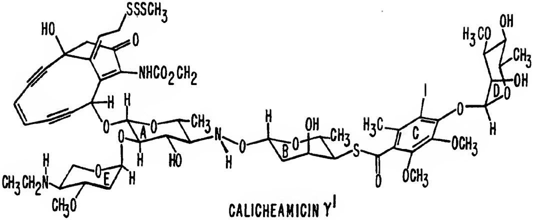
where R_1 is hydrogen, halogen, hydroxy, lower alkyl, lower alkoxy, or aryloxy; R_2 is a halogen, hydroxy, or lower alkyl; R_3 is lower alkyl, lower alkoxy, or hydrogen; and R_4 and R_5 are independently selected from lower alkyl, lower alkoxy, or hydrogen; and

detecting the ability of the glycoconjugate DNA ligand to modulate transcription of the reporter polnucleotide sequence dependent on the operably linked cis-acting transcription factor recognition sequence and predetermined transcription factor as compared to a control transcription reaction lacking the glycoconjugate DNA ligand, and;

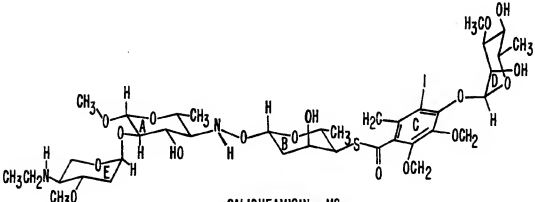
identifying a glycoconjugate DNA ligand which inhibits the predetermined transcription factor-dependent transcription of said reporter polynucleotide sequence.

22. A method of claim 21, wherein the transcription assay comprises a mammalian cell comprising a predetermined trancription factor and a reporter polynucleotide operably linked to a cis-acting transcription factor recognition sequence bound by said predetermined transcription factor.

23. A method of claim 22, wherein the mammalian cell is transfected with NFAT-SX, AP1-SX, NFkB-SX, or Oct1-SX.



CALICHEAMICIN γ[I]



CALICHEAMICIN - MG

FIG. 1.

NFAT	AaggaggaaaaaCTGTTTCAT
API	GTGACTCAGCGCG
NFKB	aggg ATTTCAC
OCTI	TGTAATATGTaaaaCAttttG
OCTI/OAP	GATCTTTGaaaaTATGTGTAATATGTaaaaCAttttG
HNFI	CAAACTGTCAAATATTAACTAAaggg AG
SPI	GGGGCGGGGC

FIG. 2.

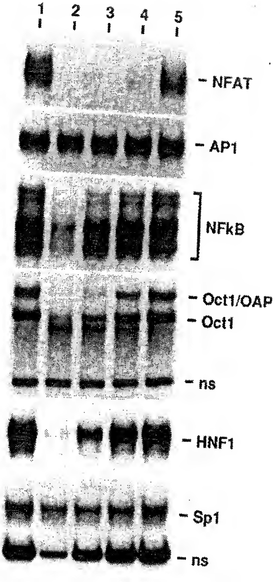


FIG. 3.

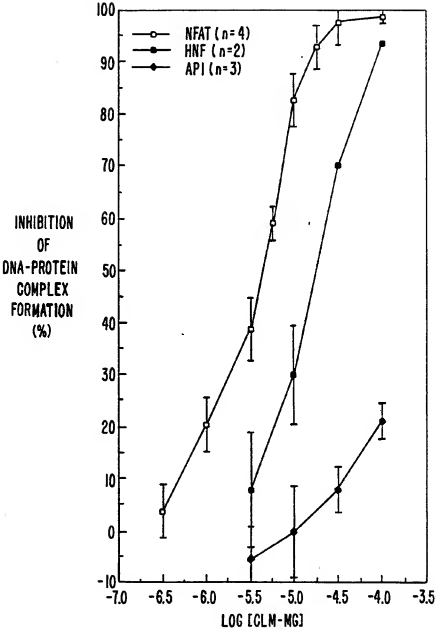

NFAT (n=4)
HNF (n=2)
API (n=3)
INHIBITION OF DNA-PROTEIN COMPLEX FORMATION (%)
100
90
80
70
60
50
40
30
20
10
0
-10
-7.0
-6.5
-6.0
-5.5
-5.0
-4.5
-4.0
-3.5
LOG [CLM-MG]

FIG. 4.

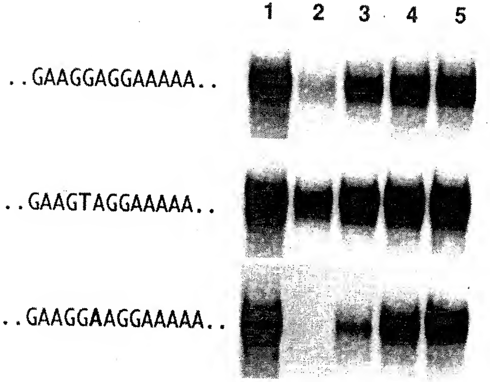


FIG. 5.

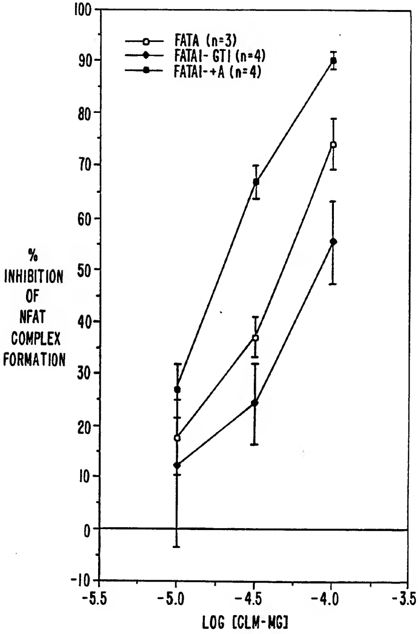

100
90
80
70
60
50
40
30
20
10
0
-10
% INHIBITION OF NFAT COMPLEX FORMATION
FATA (n=3)
FATAI- GTI (n=4)
FATAI-+A (n=4)
-5.5
-5.0
-4.5
-4.0
-3.5
LOG [CLM-MG]

FIG. 6.

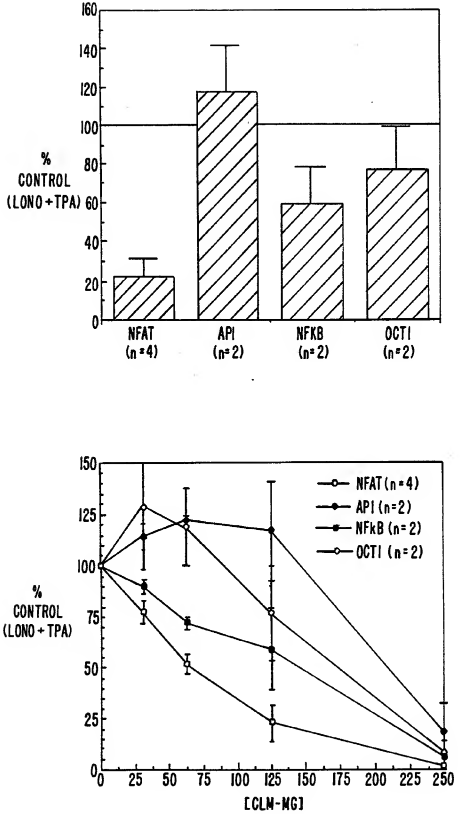

160
140
120
100
80
60
40
20
0
%
CONTROL
(LONO+TPA)
NFAT
(n=4)
API
(n=2)
NFKB
(n=2)
OCTI
(n=2)
150
125
100
75
50
25
0
%
CONTROL
(LONO+TPA)
NFAT (n=4)
API (n=2)
NFkB (n=2)
OCTI (n=2)
0 25 50 75 100 125 150 175 200 225 250
[CLM-MG]

FIG. 7.